UC San Diego

Recent Work

Title

Testing for Unit Roots with Prediction Errors

Permalink

<https://escholarship.org/uc/item/8pc6n1j8>

Author

Sanchez, Ismael

Publication Date

1998-08-01

98-21

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

TESTING FOR UNIT ROOTS WITH PREDICTION ERRORS

BY

ISMAEL SÁNCHEZ

**DISCUSSION PAPER 98-21
AUGUST 1998**

Testing for Unit Roots with Prediction Errors

Ismael Sánchez
Universidad de Alicante

July 1998

Abstract

This paper analyzes the relationship between the properties of the prediction errors of a predictor that assumes an autoregressive unit root and its optimal detection. According with this relationship, new autoregressive unit root tests are proposed based on multi-step prediction errors. It is shown that the proposed tests have optimal properties. In the simple AR(1) case, they have similar power to existing tests and very close to the Gaussian power envelope. However, in the general ARMA case, the competing tests have a high size distortion whereas the size distortion of the proposed tests is very small.

KEYWORDS: Optimal tests, predictive mean square error, unit roots.

JEL classification: C12, C22, C52.

The author is especially grateful to Daniel Peña for valuable discussions and comments. Part of this research was conducted while the author was visiting the University of California, San Diego. He is indebted to this institution. This research was supported in part by CICYT, grants PB93-0232 and PB96-0339. The usual disclaimer applies.

Correspondence address: Ismael Sánchez, Universidad de Alicante, Departamento de Fundamentos del Análisis Económico, Ap. Correos 99, 03080 Alicante (Spain). Phone: (34)965 90 36 14. Fax: (34) 965 90 38 98. Email: ismael@merlin.fae.ua.es.

1 Introduction

1.1 General considerations

This article analyzes the relationship between the prediction errors of a process with an autoregressive unit root and the optimal detection of that unit root. Following this relationship, new autoregressive unit root tests are developed based on in-sample multi-step prediction errors.

Since the work of Dickey & Fuller (1979), there has been extensive literature devoted to detecting autoregressive unit roots. This fact can be explained not only by the importance of the subject, but also by the absence of a feasible optimal test. In addition to this lack of optimality, there are some other questions related to testing for unit roots that admit different solutions, such as the estimation method, the extension to general $I(1)$ processes, the treatment of deterministic components, etc; that have helped to increase the volume of the literature. A possible criterion to classify the unit root tests literature would be the approach used to detect the unit root. The seminal and most popular approach is due to Fuller (1976) and Dickey & Fuller (1979) and is based on the asymptotic properties of ordinary least squares (OLS) estimator. Important variations of Dickey-Fuller tests are the extension to maximum likelihood estimators (Pantula *et al.*, 1994), the weighted symmetric estimator test (WS) (Park & Fuller, 1995, Fuller, 1996 p. 568), and the DF-GLS tests developed by Elliott *et al.* (1996) and Hwang & Schmidt (1996). A very important approach is the construction of tests following some optimal criteria (under normality). First works following this optimal approach are Sargan & Bhargava (1983) and Bhargava (1986). These authors extend the basis of optimal serial correlation tests to build approximate uniformly most powerful invariant tests (UMPI) and approximate locally best invariant tests (LBI). The point optimal invariant (POI) tests and LBI tests developed by Dufour & King (1991) also follow this optimal approach. These tests are based on the work of King (1980) and King & Hillier (1985). By Neyman-Pearson lemma, a sequence of unfeasible POI test (POI tests that use the true values of the root as the alternative), will provide the upper attainable power (power envelope). Dufour & King (1991) find that POI tests can have good performance if the alternative is properly chosen. According to this idea, Elliott *et al.* (1996) and Elliott (1997) build POI tests (P_T) using a fixed local-to-unity alternative. When the process follows an $AR(1)$, P_T tests have power close to the power envelope. It should be pointed out that DF-GLS tests and WS tests also have power close to the power envelope.

It could seem, since there are feasible tests that reach the power envelope, that the detection of a unit root is a solved problem. These tests, however, share an important drawback. The published performance of P_T , DF-GLS, and WS shows a high size distortion when applied to a more general $I(1)$ process. This size distortion is also present in all unit root tests that cannot handle MA components. For instance, a large positive moving average root can increase the size up to 0.60 instead of the nominal 0.05.

This article introduces a new approach to test for unit roots. The new tests are based on the prediction errors committed by a predictor that assumes a unit root. The test statistics are built by comparing the empirical prediction mean squared error (PMSE) of this predictor and the expected PMSE under the null hypothesis of a unit root. If the process is stationary, the

empirical PMSE is bounded although the predictor assumes a unit root, whereas if there is a unit root both empirical and expected PMSE are unbounded. The proposed tests have a power that is similar to the power envelope in the AR(1) case. The new tests have optimal properties. Therefore, they could also be considered members of the optimal tests class. In some situations, the application of this prediction-error approach gets to already known statistics. Two procedures to extend the proposed tests to a general ARMA case are introduced. Both procedures are based on fitting the best model and, therefore, do not need to rely on autoregressive approximations. In this more general situation, the tests still have high power with a very small size distortion. One of the proposed procedures for the ARMA case can be applied to other existing tests like P_T , reducing significantly their size distortion.

1.2 Notation and the model

Let $\{y_t\}$ be a discrete stochastic process. Let us assume that this process contains a deterministic component d_t and a pure stochastic component x_t ; namely, $y_t = d_t + x_t$. It will be assumed that the deterministic component can be a mean μ and a deterministic trend: $d_t = \mu + \delta t$. The pure stochastic part will have the following structure: $x_t = \rho x_{t-1} + u_t$, with $E(u_t) = 0$ and $E(u_t) = \sigma_u^2$. The null hypothesis of unit root corresponds with the case $\rho = 1$ and the stationary alternative with $|\rho| < 1$. Let us assume that u_t is a stationary and invertible process that admits an ARMA(p, q) representation, $\phi(B)u_t = \theta(B)a_t$, where a_t is a sequence of iid random variables with $E(a_t) = 0$ and $E(a_t^2) = \sigma^2$. The data-generating process is, therefore,

$$y_t = \mu + \delta t + x_t, \quad x_t = \rho x_{t-1} + u_t, \quad u_t = \psi(B)a_t, \quad (1.1)$$

with $\psi(B) = \phi(B)^{-1}\theta(B)$. This process can also be expressed as

$$y_t = \mu(1 - \rho) + \delta\rho + \delta(1 - \rho)t + \rho y_{t-1} + \psi(B)a_t. \quad (1.2)$$

1.3 Outline of the paper

This article shows the links between the optimal detection of unit roots and the properties of the PMSE and introduces new tests. Section 2 shows the behavior of the PMSE of a random walk predictor when the true process is stationary. The boundness of this PMSE is the motivation of the proposed test statistics. Section 3 defines and derives the most powerful tests in a neighborhood of the locality and show that they also have a prediction-error interpretation. The resulting optimal test statistics are functions of the statistics proposed in section 2. Section 4 uses the prediction-error interpretation to modify the optimal statistics and obtain power similar to the power envelope. Section 5 compares the proposed tests with some other existing in the literature through a Monte Carlo experiment in the AR(1) case. Section 6 extends the tests to the general I(1) case and compares them with other tests. This section analyzes two efficient procedures that allow for the modellization of MA components with very small distortion. Section 7 concludes.

2 Tests based on the ratio of observed and expected PMSE.

This section shows the basic test statistics that compare the empirical PMSE and the expected one under the null hypothesis. Let us suppose $\{y_t\}$ follows a pure AR(1) process with no deterministic components. The observation y_{T+h} can be expressed as $y_{T+h} = \rho^h y_T + \sum_{j=0}^{h-1} \rho^j a_{T+h-j}$. Suppose now that it is believed that the process has a unit root and, therefore, the predictor

$$\tilde{y}_{T+h} = y_T \quad (2.3)$$

is used. Let us compare the PMSE of this predictor with the one it is expected given that a unit root is assumed. First, note that the predictor is biased, since $E(y_{T+h}|y_T) = \rho^h y_T$, and therefore, its PMSE is

$$E(y_{T+h} - y_T)^2 = \sigma^2 \left\{ \sum_{j=0}^{h-1} \rho^{2j} + \frac{(1 - \rho^h)^2}{1 - \rho^2} \right\} = 2\sigma^2 \frac{1 - \rho^h}{1 - \rho^2}, \quad (2.4)$$

where it has been used that $E(y_T^2) = \sigma^2/(1 - \rho^2)$. The expected PMSE is, however, the one associated to a random walk. In order to compare the real and expected PMSE, the expression (2.4) can be rewritten as $E(y_{T+h} - y_T)^2 = \sigma^2 h d(h, \rho)$, where

$$d(h, \rho) = \frac{E(y_{T+h} - y_T)^2}{E(y_{T+h} - y_T | \rho = 1)^2} = 2 \frac{1 - \rho^h}{h(1 - \rho^2)}. \quad (2.5)$$

Note that, for forecasting horizon $h = 1$, $d(1, \rho) = 2/(1 + \rho) > 1$ and, therefore, the actual PMSE is larger than σ^2 . For $h = 2$, it holds that $d(2, \rho) = 1$, and the actual PMSE is identical to the expected. For $h > 2$, $d(h, \rho) < 1$ and the real PMSE is always lower than expected. Similarly, for a given $h > 2$, the ratio of the real PMSE and the expected one goes to zero as ρ moves away from unity. Therefore, the ratio of the real PMSE and the expected one moves to zero as the process moves away from the unit root. Furthermore, this effect increases with the horizon. The proposed tests are based on this property.

Let us define $e_{i|j}$ as the empirical prediction error of forecasting y_i from $t = j$ ($i > j$) under the null hypothesis of random walk. Then, $e_{i|j} = y_i - y_j$. The test statistics are built by comparing this empirical prediction errors and the expected ones under the null. First, let us base the test statistics on the cumulated sum of prediction square errors from $h = 1$ to $h = T - 1$ given by

$$\sum_{t=2}^T e_{t|1}^2 = \sum_{t=2}^T (y_t - y_1)^2. \quad (2.6)$$

The expected value of this statistic under the null hypothesis of a unit root is:

$$E \left(\sum_{t=2}^T e_{t|1}^2 \right) = \sigma^2 \{1 + 2 + \dots + (T - 1)\} = \sigma^2 \frac{T(T - 1)}{2}. \quad (2.7)$$

Since σ^2 is not known, the following consistent estimator, under the null, can be used:

$$\hat{\sigma}^2 = \frac{\sum_{t=2}^T (y_t - y_{t-1})^2}{T - 1}. \quad (2.8)$$

The test statistic is obtained by dividing the cumulated sum (2.6) and its expected value under the null (2.7) using the estimation (2.8). The constant can be dropped since it does not affect the test. The proposed statistic is the following:

$$C_1^\mu = \frac{\sum_{t=2}^T (y_t - y_1)^2}{T \sum_{t=2}^T (y_t - y_{t-1})^2}, \quad (2.9)$$

where C denotes cumulated and the subindex shows the origin of the predictions. The test statistic is invariant to transformation of the type $y_t \rightarrow ay_t + b$, with a, b constants and it will not be affected by the mean value of the series. It can be verified that, under the null hypothesis of a unit root, $C_1^\mu = O_p(1)$. Under the alternative, the numerator is, however, of lower magnitude order than in the previous case since $T^{-2} \sum_{t=2}^T y_t^2 = O_p(T^{-1})$. Therefore, in the stationary case, $C_1^\mu \xrightarrow{p} 0$. A unit root test against stationary alternative will have the rejection region $C_1^\mu < c_1^\mu$.

Let us now extend this test to the case of a null hypothesis of a random walk with drift, $y_t = \delta + y_{t-1} + a_t$. The prediction errors are $e_{t|1} = y_t - y_1 - (t-1)\delta$. An efficient estimator of δ under the null can be obtained from the regression $\Delta y_t = \delta + a_t$. Therefore,

$$\hat{\delta} = \frac{\sum_{t=2}^T (y_t - y_{t-1})}{T-1} = \frac{y_T - y_1}{T-1}. \quad (2.10)$$

A consistent estimator of σ^2 , under the null, is

$$\hat{\sigma}^2 = \frac{\sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}{T-1}.$$

This leads to the use of the following test statistic:

$$C_1^\tau = \frac{\sum_{t=2}^T \{y_t - y_1 - \hat{\delta}(t-1)\}^2}{T \sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}. \quad (2.11)$$

Similar to C_1^μ , the rejection region of a unit root test with this statistic is $C_1^\tau < c_1^\tau$.

Both statistics, C_1^μ and C_1^τ , have the feature that their first two terms of the numerator behave, on average, differently than the others, as previously indicated with the distance d in (2.5). A solution to overcome this problem would be to consider only prediction errors corresponding to horizon $h > 1$. This modification, however, has not improved the performance of the statistics. A second feature of the proposed statistics is their high dependence on the first observation since all prediction errors are evaluated from y_1 . Two possible alternatives to diminish such dependence are presented. First, C_1^μ and C_1^τ use only one prediction error at each horizon, whereas many more prediction errors can be considered. There are $(T-1)$ prediction errors at $h = 1$; namely, $(y_2 - y_1), (y_3 - y_2), \dots, (y_T - y_{T-1})$; and, in general, $(T-k)$ at horizon $h = k$. The following statistics uses all the available information in the sample by averaging all prediction errors of the same horizon:

$$C_A^\mu = \frac{\sum_{t=1}^{T-1} \sum_{s=t+1}^T \{T - (s-t)\}^{-1} (y_s - y_t)^2}{T \sum_{t=2}^T (y_t - y_{t-1})^2}, \quad (2.12)$$

$$C_A^\tau = \frac{\sum_{t=1}^{T-1} \sum_{s=t+1}^T \{T - (s-t)\}^{-1} \{y_s - y_t - \hat{\delta}(s-t)\}^2}{T \sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}, \quad (2.13)$$

where A stands for average. Another possible solution to diminish the dependence on the first observation is to consider also the series in time reverse order. The use of the time reversal series is justified by the time-reversibility of stationary processes (Box & Jenkins, 1976, p. 197). The statistic that, at each observation, average the prediction error from both extremes of the series is

$$C_{1T}^\mu = \frac{\sum_{t=2}^T (y_t - y_1)^2 + \sum_{t=1}^{T-1} (y_t - y_T)^2}{2T \sum_{t=2}^T (y_t - y_{t-1})^2}, \quad (2.14)$$

where the subindex $1T$ denotes the origin of the predictions. If the process is stationary both statistics, C_{1T}^μ and C_1^μ , will have similar behavior, but their properties under the null are different. Several authors have also used the time-reversibility property to modify the statistics proposed by Dickey-Fuller and obtained more powerful tests (Sen & Dickey, 1987, Pantula *et al.*, 1994, Leybourne, 1995, Park & Fuller, 1995, among others). It can be checked that $C_{1T}^\tau = C_1^\tau$.

The statistics C_1^μ and C_1^τ are already in the literature using different approaches. They correspond to $(TN_1)^{-1}$ and $(TN_2)^{-1}$ respectively, where N_1 and N_2 are in Bhargava (1986). Bhargava (1986) shows that, if the Anderson approximation is used for the inverse of the covariance of the process and the first observation is extracted from its conditional distribution, these statistics lead to approximate LBI tests under normality. This statistics are also derived in Tanaka (1997), with the same assumption about the first observation, by taking the second derivative of the likelihood function evaluated at the maximum likelihood estimator under the null.

The following theorem shows the asymptotic distribution of the statistics C_1^μ , C_{1T}^μ , and C_A^μ when the process is $y_t = y_{t-1} + u_t$, $u_t = \psi(B)a_t$, and that of C_1^τ and C_A^τ when the process is $y_t = \delta + y_{t-1} + u_t$, $u_t = \psi(B)a_t$. The notation $\xrightarrow{d}$ denotes weak convergence of the associated probability measures as $T \rightarrow \infty$. The proof can be found in the appendix.

Theorem 1 *Let y_t be the process (1.2) with $\rho = 1$. Then,*

$$C_1^\mu \xrightarrow{d} \kappa^{-2} \int_0^1 W^2(r) dr, \quad (2.15)$$

$$C_1^\tau \xrightarrow{d} \kappa^{-2} \left\{ \int_0^1 W^2(r) dr + \frac{1}{3} W^2(1) - 2W(1) \int_0^1 r W^2(r) dr \right\}, \quad (2.16)$$

$$C_{1T}^\mu \xrightarrow{d} \kappa^{-2} \left\{ \int_0^1 W^2(r) dr + \frac{1}{2} W^2(1) - W(1) \int_0^1 W(r) dr \right\}, \quad (2.17)$$

$$C_A^\mu \xrightarrow{d} \kappa^{-2} \int_0^1 \int_0^r \{1 - (r - \lambda)\}^{-1} W^2(r - \lambda) d\lambda dr, \quad (2.18)$$

$$C_A^\tau \xrightarrow{d} \kappa^{-2} \left[\int_0^1 \int_0^r \{1 - (r - \lambda)\}^{-1} W^2(r - \lambda) d\lambda dr + \frac{1}{3} W^2(1) - 2W(1) \int_0^1 \int_0^r \{1 - (r - \lambda)\}^{-1} (r - \lambda) W(r - \lambda) d\lambda dr \right], \quad (2.19)$$

where

$$\kappa^2 = \frac{\sigma_u^2}{\omega^2}, \quad \sigma_u^2 = E(u^2), \quad \omega^2 = \sigma^2 \psi(1)^2.$$

In the general I(1) case, with $u_t = \psi(B)a_t$, the asymptotic distributions depends on the parameters ψ_j , σ_u^2 , and σ^2 . Section 6 shows modifications to allow, in the general I(1) case, to use the

same critical values than in the AR(1) case.

3 Optimal tests and prediction errors

This section shows asymptotic tests following optimal criteria under normality of the disturbances. It is found that the resulting statistics are the same as the statistics proposed in the last section or functions of them. The proposed tests are based on the theory developed by King (1980), King & Hillier (1985) and Dufour & King (1991), among others, to build POI and LBI tests. The interest is in finding tests that are most powerful in a neighborhood of the locality (MPNL). In order to obtain this property, a Taylor expansion of the POI test statistics is used around the null hypothesis up to terms of second order. By Neyman-Pearson lemma (see King & Hillier, 1985, and Ferguson, 1967) this approach will give a power function with maximum first and second derivatives at $\rho = 1$. A third order Taylor expansion has also been analyzed, but the inclusion of the third term does not produce any improvement. This section develops tests for the AR(1) case. Section 6 extend the proposed tests to a more general case.

Let us assume, first, that the model is as (1.1) with $\delta = 0$ and $\psi(B) = 1$. Let us also assume that u_t is normally distributed. Two different assumptions about x_0 will be considered. First, it is considered that x_0 is fixed and, therefore, y_1 is extracted from its conditional distribution. Second, x_0 is considered random with zero mean and variance $\sigma^2/(1 - \rho^2)$. Then, y_1 is extracted from its unconditional distribution. Under the first assumption of x_0 fixed, the process x_t is non-stationary. This assumption will be denoted as the non-stationary case. The covariance matrix of the $T \times 1$ vector $\mathbf{x} = (x_1, \dots, x_T)'$ is $\sigma^2 \mathbf{\Omega}_N(\rho)$, where

$$\mathbf{\Omega}_N^{-1}(\rho) = \mathbf{I}_T - \rho(\mathbf{L}_T + \mathbf{L}_T') + \rho^2 \mathbf{L}_T' \mathbf{L}_T, \quad (3.20)$$

with $\mathbf{I}_T$ a $T \times T$ identity matrix and $\mathbf{L}_T$ a $T \times T$ matrix with 1s on the diagonal immediately below the main diagonal and 0s elsewhere. It will be denoted $\mathbf{\Omega}_1 = \mathbf{\Omega}_N(1)$. This non-stationary model is the same as the Dufour & King (1991) model with $d_1 = 1$. Following these authors, the POI test of $\rho = 1$ against $\rho = \rho_0$ rejects the null hypothesis for small values of

$$S_N(\rho_0) = \frac{\hat{\mathbf{x}}_N' \mathbf{\Omega}_N^{-1}(\rho_0) \hat{\mathbf{x}}_N}{\hat{\mathbf{x}}_1' \mathbf{\Omega}_1^{-1} \hat{\mathbf{x}}_1}, \quad (3.21)$$

where $\hat{\mathbf{x}}_N$ and $\hat{\mathbf{x}}_1$ are generalized least squares (GLS) residual vectors corresponding to covariance matrix $\mathbf{\Omega}_N(\rho_0)$ and $\mathbf{\Omega}_1$, respectively. Let us consider now the case that x_0 is random with zero mean and variance $\sigma^2/(1 - \rho^2)$. In this case, x_t is covariance stationary. This assumption is denoted as the stationary case. The covariance matrix of the vector $\mathbf{x}$ is $\sigma^2 \mathbf{\Omega}_S(\rho)$, where

$$\mathbf{\Omega}_S^{-1}(\rho) = \mathbf{\Omega}_N^{-1}(\rho) - \rho^2 \mathbf{e} \mathbf{e}', \quad (3.22)$$

with $\mathbf{e} = (1, 0, \dots, 0)' (T \times 1)$. This model corresponds to the Dufour & King (1991) model with $d_1 = (1 - \rho^2)^{-1/2}$. Following these authors, the POI tests of $\rho = 1$ against $\rho = \rho_0$ rejects for small values of

$$S_S(\rho_0) = \frac{\hat{\mathbf{x}}_S' \mathbf{\Omega}_S^{-1}(\rho_0) \hat{\mathbf{x}}_S}{\hat{\mathbf{x}}_1' \mathbf{\Omega}_1^{-1} \hat{\mathbf{x}}_1}, \quad (3.23)$$

where $\hat{\mathbf{x}}_S$ is the generalized least squares (GLS) residual vector corresponding to covariance matrix $\boldsymbol{\Omega}_S(\rho_0)$. The non-stationary case is a reasonable assumption when the time series has its origin at x_0 , whereas the stationary case reflects the situation of a time series with the origin in a very far past point. It seems reasonable to think that most real series will adapt better to the stationary case. The following propositions show the Taylor series expansion of (3.21) and (3.23) around $\rho = 1$. See proof in the appendix.

Proposition 3.1 *Let $y_t = \mu + x_t$, where x_t follows an $AR(1)$ process with x_0 fixed. Then,*

$$S_N(\rho) = \frac{\hat{\mathbf{x}}_N' \boldsymbol{\Omega}_N^{-1}(\rho) \hat{\mathbf{x}}_N}{\hat{\mathbf{x}}_1' \boldsymbol{\Omega}_1^{-1} \hat{\mathbf{x}}_1} = 1 + (\rho - 1)S_N'(1) + \frac{(\rho - 1)^2}{2}S_N''(1) + O\{(\rho - 1)^3\},$$

where

$$\begin{aligned} S_N'(1) &= \left. \frac{\partial S_N(\rho)}{\partial \rho} \right|_{\rho=1} = 1 - \frac{(y_T - y_1)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2}, \\ S_N''(1) &= \left. \frac{\partial^2 S_N(\rho)}{\partial \rho^2} \right|_{\rho=1} = 2 \frac{\sum_{t=2}^T (y_t - y_1)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2} - 4 \frac{(y_T - y_1)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2}. \end{aligned}$$

Proposition 3.2 *Let $y_t = \mu + x_t$, where x_t follows an $AR(1)$ process with x_0 being a random variable with zero mean and variance $\sigma^2/(1 - \rho^2)$. Then,*

$$S_S(\rho) = \frac{\hat{\mathbf{x}}_S' \boldsymbol{\Omega}_S^{-1}(\rho) \hat{\mathbf{x}}_S}{\hat{\mathbf{x}}_1' \boldsymbol{\Omega}_1^{-1} \hat{\mathbf{x}}_1} = 1 + (\rho - 1)S_S'(1) + \frac{(\rho - 1)^2}{2}S_S''(1) + O\{(\rho - 1)^3\},$$

where

$$\begin{aligned} S_S'(1) &= \left. \frac{\partial S_S(\rho)}{\partial \rho} \right|_{\rho=1} = 1 - \frac{1}{2} \frac{(y_T - y_1)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2}, \\ S_S''(1) &= \left. \frac{\partial^2 S_S(\rho)}{\partial \rho^2} \right|_{\rho=1} = \frac{\sum_{t=2}^{T-1} (y_t - y_1)^2 + \sum_{t=2}^{T-1} (y_t - y_T)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2} - 2 \frac{\sum_{t=2}^T (y_t - y_1)(y_T - y_t)}{\sum_{t=2}^T (y_t - y_{t-1})^2}. \end{aligned}$$

These Taylor expansions suggest three different kinds of test statistics. At first approximation, tests statistics can be built using only the the first derivatives of the expansions. These derivatives are identical in both cases. Following Dufour & King (1991), these statistics lead to a LBI test and rejects the null against a stationary alternative for small values of

$$E^\mu = \frac{(y_T - y_1)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2}, \quad (3.24)$$

where E stands for the extreme points of the series. This statistic has previously been obtained by Nabeya & Tanaka (1990) and Tanaka (1996) for the non-stationary case. It should be noted that the approach followed by these authors is only valid under the non-stationary case assumption. This statistic also has an interpretation in terms of prediction errors. E^μ is the ratio between the empirical prediction square error of the last observation from the origin of the series, assuming a unit root, and the expected one. When $\mu = 0$ (or if μ is known and the series

is previously demeaned with that value), and following the same arguments as in the proof of these propositions, the LBI is:

$$E_S^0 = \frac{y_T^2 + y_1^2}{\sum_{t=2}^T (y_t - y_{t-1})^2}, \quad (3.25)$$

for the stationary case. In the non-stationary case, the LBI is:

$$E_N^0 = \frac{y_T^2}{\sum_{t=2}^T (y_t - y_{t-1})^2}. \quad (3.26)$$

This last statistic, E_N^0 , has been previously obtained by Stock (1994), Nabeya & Tanaka (1990), and Tanaka (1996).

Tests based on these statistics have optimal power only when the alternative is very close to the null (let say $\rho = 0.999..$). The power is very low for alternatives of practical interest. This drawback suggests a second class of statistics using only the second derivatives of the Taylor expansions. In order to obtain random variables with asymptotic distributions it is necessary to divide this second derivatives by T . This correction makes the second terms both in $S_N''(1)$ and $S_S''(1)$ converge to zero in probability under the null hypothesis of unit root. In the non-stationary case, it verifies that

$$\frac{(y_T - y_1)^2}{T \sum_{t=2}^T (y_t - y_{t-1})^2} = \frac{(y_T - y_1)^2}{T^2 \hat{\sigma}_u^2} \xrightarrow{p} 0.$$

In the stationary case, it holds that

$$\frac{\sum_{t=2}^T (y_t - y_1)(y_T - y_t)}{T \sum_{t=2}^T (y_t - y_{t-1})^2} = \frac{\sum_{t=2}^T t \hat{\gamma}_t}{T \hat{\sigma}_u^2} \xrightarrow{p} 0, \quad (3.27)$$

where $\hat{\gamma}_t = \sum_{j=1}^{T-t} u_j u_{j+t} / T$. Note that (3.27) still holds if u_t is a general stationary ARMA model. With these results, the second derivatives are asymptotically equivalent to two times C_{1T}^μ in the non-stationary case and C_{1T}^μ in the stationary case, which were derived in the last section using the prediction-error interpretation.

It seems that the stationary case uses time-reversibility to obtain optimal properties. This is not only seen in the asymptotic equivalence of $S_S(\rho)''$ and C_{1T}^μ , but also (3.24) and (3.25) admit this time-reversibility interpretation. When $\mu = 0$, equation (3.26) shows that the LBI, in the non-stationary case, is based on the last observation. If the time reversal series is also considered, the LBI test should use both extremes as (3.25) does. If μ is unknown, the time-reversibility can also explain the fact that the LBI statistic (3.24) is the same both in the stationary and non-stationary case. This optimal property of the time-reversibility can also help to understand why the modifications of Dickey-Fuller tests based on using also the series in reverse order improve their performance.

If $\mu = 0$, the second derivative leads to the statistics $C_1^0 = (\sum_{t=1}^{T-1} y_t^2) / \{T \sum_{t=2}^T (y_t - y_{t-1})^2\}$ and $C_{1T}^0 = (\sum_{t=1}^{T-1} y_t^2 + \sum_{t=2}^T y_t^2) / \{2T \sum_{t=2}^T (y_t - y_{t-1})^2\}$ for the non-stationary and stationary case, respectively.

A third type of tests using this optimal approach can be obtained by using both first and second derivatives as they appear in the Taylor expansion. These tests will be denoted as more powerful in a neighborhood of the locality (MPNL). It will be convenient to parameterize the alternative as $\rho_c = 1 - c/T$, with $c > 0$. It can be checked that, in the non-stationary case, the asymptotic MPNL test will reject the null hypothesis for small values of

$$\begin{aligned} EC_1^0 &= E_N^0 + cC_1^0, \\ EC_1^\mu &= E^\mu + cC_1^\mu; \end{aligned} \tag{3.28}$$

whereas for the stationary case,

$$\begin{aligned} EC_{1T}^0 &= E_S^0 + cC_{1T}^0, \\ EC_{1T}^\mu &= E^\mu + cC_{1T}^\mu. \end{aligned} \tag{3.29}$$

The following proposition extends the previous results to the deterministic trend case. Since the proposition holds for both stationary and non-stationary case, differences in notation will be omitted. The proof is in the appendix. This proposition extends the results of Nabeya & Tanaka (1990) and Tanaka (1996) to the stationary case. It should be noted that the methodology exposed in Tanaka (1996) to derive optimal tests in the non-stationary case is not easily applicable to the stationary case, since it would deal with non-invertible matrices.

Proposition 3.3 *Let $y_t = \mu + \delta t + x_t$, where x_t follows an $AR(1)$ process with x_0 either fixed or following a random variable with zero mean and variance $\sigma^2/(1 - \rho^2)$. Then,*

$$S(\rho) = \frac{\hat{\mathbf{x}}' \boldsymbol{\Omega}^{-1}(\rho) \hat{\mathbf{x}}}{\hat{\mathbf{x}}_1' \boldsymbol{\Omega}_1^{-1} \hat{\mathbf{x}}_1} = 1 + (\rho - 1)S'(1) + \frac{(\rho - 1)^2}{2}S''(1) + O\{(\rho - 1)^3\},$$

where

$$\begin{aligned} S'(1) &= \left. \frac{\partial S(\rho)}{\partial \rho} \right|_{\rho=1} = 1, \\ S''(1) &= \left. \frac{\partial^2 S(\rho)}{\partial \rho^2} \right|_{\rho=1} = 2 \frac{\sum_{t=2}^T \{y_t - y_1 - \hat{\delta}(t-1)\}^2}{\sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}. \end{aligned}$$

Since the first derivative is a constant, a LBI test would be of no practical interest. The second derivative can be used to build an asymptotic LBIU test for both the stationary and non-stationary cases. It can be checked that dividing the second derivative by T , the asymptotic LBIU uses the statistics C_1^τ , derived in section 2 using the prediction-error interpretation. This result is also compatible with the time-reversibility property used in the stationary case since, as seen in section 2, $C_1^\tau = C_{1T}^\tau$. Therefore, C_1^τ is the asymptotic MPNL test when there is a deterministic trend. The asymptotic distributions of the statistics E^μ , EC_1^μ , and EC_{1T}^μ , for the general case of $u_t = \psi(B)a_t$, are shown in the following proposition. The asymptotic distributions for the pure stochastic case are the same. The proof is a straightforward application of the proof of theorem 1.

Proposition 3.4 *Let y_t be the process (1.2) with $\rho = 1$ and $\delta = 0$. Then,*

$$E^\mu \xrightarrow{d} \kappa^{-2} W^2(1), \quad (3.30)$$

$$EC_1^\mu \xrightarrow{d} \kappa^{-2} \left\{ W^2(1) + c \int_0^1 W^2(r) dr \right\}, \quad (3.31)$$

$$EC_{1T}^\mu \xrightarrow{d} \kappa^{-2} \left[W^2(1) + c \left\{ \int_0^1 W^2(r) dr + \frac{1}{2} W^2(1) - W(1) \int_0^1 W(r) dr \right\} \right]. \quad (3.32)$$

The asymptotic distributions depends on the parameter κ^2 . Section 6 shows modifications of these statistics to avoid this dependency. The asymptotic distribution of EC_1^μ and EC_{1T}^μ also depends on c . In the non-stationary case (statistic EC_1^μ), computations show that the value of c that provides better power, improving the performance of C_1^μ , is $c = 7$. In the stationary case, however, the statistic EC_{1T}^μ does not improve the performance of C_{1T}^μ .

4 Locally biased predictors

This section modifies the proposed statistics in order to obtain better power. The modified statistics will use predictors built under a fixed local alternative, instead of under the null of a unit root. The information regarding the fixed local alternative $\rho_c = 1 - c/T$ has already been used in the definition of MPNL tests. This information is summarized in the parameter c of the statistic (3.28) and (3.29). In this section, the prediction-error interpretation will be used to extend the information of the local alternative also in the predictors. The idea is related with the modification of the Dickey-Fuller tests proposed by Elliott *et al.* (1996) and Hwang & Schmidt (1996) (DF-GLS). These authors eliminate the deterministic components assuming that, under the alternative, the process is nearly non-stationary with $\rho_c = 1 - c/T$, $c > 0$. This leads to estimate the deterministic parameters by GLS under such alternative. As a result, these authors obtain that DF-GLS tests have higher power than the original Dickey-Fuller tests.

Let us suppose that the process follows the model (1.1) with $\psi(B) = 1$. Under the null hypothesis of a unit root, the h -steps ahead prediction is $\hat{y}_{t+h} = \delta h + y_t$. Test statistics proposed in last sections use this predictor. It will now be considered that the predictor assumes that the process is nearly non-stationary. Therefore, the new locally biased predictor is

$$\hat{y}_{t+h} = \mu(1 - \rho_c^h) + \delta t(1 - \rho_c^h) + \delta h + \rho_c^h y_t, \quad (4.33)$$

where the parameters μ and δ are optimally estimated by GLS. All previously seen statistics improve their power with this approach, with the exception of the statistics C_A^μ and C_A^τ . The best performance is obtained by the modifications of the MPNL tests derived in last section. Therefore, attention will be restricted to these tests. Let us denote $\hat{\mu}_S$ and $\hat{\delta}_S$ to the estimators of the deterministic components when the matrix $\mathbf{\Omega}_S^{-1}(\rho_c)$ is used, and $\hat{\mu}_N$ and $\hat{\delta}_N$ when the estimation is made with $\mathbf{\Omega}_N^{-1}(\rho_c)$. Let us also denote $y_{t|S}^\mu = y_t - \hat{\mu}_S$ and $y_{t|S}^\tau = y_t - \hat{\mu}_S - \hat{\delta}_S t$ in the stationary case. Similarly, $y_{t|N}^\mu$ and $y_{t|N}^\tau$ will be used in the non-stationary case. In the non-stationary case the modified MPNL test statistics are:

$$B_N^\mu = \frac{(y_{T|N}^\mu - \rho_c^{T-1} y_{1|N}^\mu)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2} - c \frac{\sum_{t=2}^T (y_{t|N}^\mu - \rho_c^{t-1} y_{1|N}^\mu)^2}{T \sum_{t=2}^T (y_t - y_{t-1})^2} \quad (4.34)$$

and

$$B_N^\tau = \frac{\sum_{t=2}^T (y_{t|N}^\tau - \rho_c^{t-1} y_{1|N}^\tau)^2}{T \sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}, \quad (4.35)$$

where B stands for biased predictor. Likewise, for the stationary case, series can be detrended by GLS using the matrix $\boldsymbol{\Omega}_S^{-1}(\rho_c)$. Computations show, however, that better results are obtained if, in the non-zero mean case, the matrix $\boldsymbol{\Omega}_N^{-1}(\rho_c)$ is used instead. Let us denote z_t to the time-reversal process ($z_1 = y_T, \dots, z_T = y_1$). The modified MPNL test statistics are

$$B_S^\mu = \frac{(y_{T|N}^\mu - \rho_c^{T-1} y_{1|N}^\mu)^2 + (z_{T|N}^\mu - \rho_c^{T-1} z_{1|N}^\mu)^2}{2 \sum_{t=2}^T (y_t - y_{t-1})^2} - c \frac{\sum_{t=2}^T (y_{t|N}^\mu - \rho_c^{t-1} y_{1|N}^\mu)^2 + \sum_{t=2}^T (z_{t|N}^\mu - \rho_c^{t-1} z_{1|N}^\mu)^2}{2T \sum_{t=2}^T (y_t - y_{t-1})^2} \quad (4.36)$$

and

$$B_S^\tau = \frac{\sum_{t=2}^T (y_{t|S}^\tau - \rho_c^{t-1} y_{1|S}^\tau)^2}{T \sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}. \quad (4.37)$$

Let us denote as C to the set of test statistics developed in sections 2 and 3, and $C(c)$ to the corresponding modified statistics following the locally biased predictor approach, including the B test statistics. The following proposition shows that both types of test statistics have the same asymptotic distribution. See proof in appendix.

Proposition 4.1 *Let y_t be the process (1.2) with $\rho = 1$ and let $\rho_c = 1 - c/T$. Then,*

$$C - C(c) \xrightarrow{d} 0.$$

5 Finite sample performance. AR(1) case

This section shows a simulation exercise to evaluate the empirical power of the proposed tests and compare them with some others that appear in the literature. Elliott *et al.* (1996) showed that currently used tests, apart from the LBI test, have all power function very close to the power envelope in absence of deterministic components. Therefore, only processes with some deterministic component will be considered. The Monte Carlo experiment has been made for sample sizes $T = 50, 100$, but only $T = 100$ is reported. Conclusions are identical in both sample sizes.

Critical values for the proposed tests have been obtained through Monte Carlo simulations with 100,000 replications and can be found in Table 1. Since asymptotic distributions do not depend on δ , the value $\delta = 0$ has been used. The simulated process is $y_t = \rho y_{t-1} + a_t$, with $a_t \sim N(0, 1)$ and $y_1 = a_1$. It is reported the simulations of the basic statistics derived in section 2; namely, C_1 , C_{1T} and C_A ; and the B tests of section 4 with better performance. For the non-stationary case, the best B statistics are $B_N^\mu(c = 7.5)$ and $B_N^\tau(c = 11)$. In the stationary case, the B statistics with better performance are $B_S^\mu(c = 11.5)$ and $B_S^\tau(c = 11)$. More detailed critical values of these B tests can be found in the appendix.

The tests that will be used for comparison are the asymptotic bias $T(\hat{\rho} - 1)$ and $\hat{\tau}$ proposed by Dickey & Fuller (1979), Sargan-Bhargava tests (SB) (Sargan & Bhargava, 1983, Bhargava, 1986); the P_T tests and DF-GLS ($\hat{\tau}_{GLS}$) tests proposed by Elliott *et al.* (1996) and Elliott (1997); and the weighted symmetric estimator tests ($\hat{\tau}_{WS}$) proposed by Park & Fuller (1995) (see also Pantula *et al.*, 1994 and Fuller, 1997).

The Sargan-Bhargava tests that are included in this comparison are:

$$\begin{aligned} \text{SB}^\mu &= \frac{\sum_{t=1}^T (y_t - \bar{y})^2}{T \sum_{t=2}^T (y_t - y_{t-1})^2}, \\ \text{SB}^\tau &= \frac{\sum_{t=1}^T \{y_t - \bar{y} - \hat{\delta}(t-1) + (y_T - y_1)/2\}^2}{T \sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2}, \end{aligned}$$

where $\bar{y} = T^{-1} \sum_{t=1}^T y_t$. Note that $\text{SB}^\mu = T R_1^{-1}$ and $\text{SB}^\tau = T R_2^{-1}$, where R_1 and R_2 are in Bhargava (1986). The P_T tests are based on the point optimal tests proposed by Dufour & King (1991) evaluated at $\rho_c = 1 - c/T$. In the stationary case (using the matrix $\boldsymbol{\Omega}_S^{-1}(\rho_c)$), Elliott (1997) suggests the value $c = 10$ for both P_T and $\hat{\tau}_{GLS}$. In the non-stationary case (detrending with the matrix $\boldsymbol{\Omega}_N^{-1}(\rho_c)$), Elliott *et al.* (1996) suggest, for these two tests, $c = 7$ when the alternative has a non-zero mean, and $c = 13.5$ for the deterministic trend. The scaled unfeasible point optimal test of Dufour & King (1991) (DK), using the true parameter ρ as alternative, is also included as asymptotic power bound.

Critical values for Dickey-Fuller tests are in Fuller (1976). Critical values of $\hat{\tau}_{GLS}^\tau$, P_T^μ , and P_T^τ are in Elliott *et al.* (1996) for the non-stationary case. Critical values for the remaining tests have been obtained through Monte Carlo and can be found in Table 2. This table has been created following a similar experiment as in Table 1. Critical values for $\hat{\tau}_{GLS}^\mu$ are not reported by their authors since they use the asymptotic values of Dickey-Fuller tests. In order to ease the comparison between tests and use the same size, new critical values has been obtained by Monte Carlo. These simulated values coincide with those reported by Pantula *et al.* (1994).

In order to evaluate the power, experiments have been performed for the stationary and non-stationary cases. The number of replication is also 100,000. In the stationary case, the simulated process is $y_t = \rho y_{t-1} + a_t$ with $a_t \sim N(0, 1)$ and $y_1 = a_1 / \sqrt{1 - \rho^2}$, whereas in the nonstationary case $y_1 = a_1$. Tables 3 and 4 show the empirical power in the stationary case, whereas Tables 5 and 6 show the empirical power in the non-stationary case.

Table 1: Critical values at 5%. $T = 100$.

C_1^μ	0.0581
C_{1T}^μ	0.0722
C_A^μ	0.0819
C_1^τ	0.0380
C_A^τ	0.0562
$B_N^\mu(c = 7.5)$	0.4408
$B_N^\tau(c = 9)$	0.0298
$B_S^\mu(c = 11.5)$	0.5873
$B_S^\tau(c = 11)$	0.0264

Table 2: Critical values at 5%. $T = 100$.

SB^μ	0.038
SB^τ	0.029
$P_T^\mu(c = 10)$ (stationary case)	4.713
$P_T^\tau(c = 10)$ (stationary case)	2.942
$\hat{\tau}_{GLS}^\mu(c = 10)$ (stationary case)	-2.76
$\hat{\tau}_{GLS}^\tau(c = 10)$ (stationary case)	-3.22
$\hat{\tau}_{GLS}^\mu(c = 7)$ (non-stationary case)	-2.14
$\hat{\tau}_{WS}^\mu$	-2.56
$\hat{\tau}_{WS}^\tau$	-3.28

Table 3: Empirical power at size 5%. $T = 100$. Stationary case. $d_t = \mu$. 100,000 rep.

Test	ρ				
	0.97	0.95	0.90	0.80	0.70
C_1^μ	0.11	0.18	0.42	0.75	0.89
C_{1T}^μ	0.12	0.20	0.47	0.84	0.96
C_A^μ	0.11	0.19	0.51	0.96	1.00
B_S^μ	0.12	0.20	0.53	0.96	1.00
$\hat{\tau}_{WS}^\mu$	0.12	0.20	0.52	0.97	1.00
P_T^μ	0.12	0.20	0.52	0.97	1.00
$\hat{\tau}_{GLS}^\mu$	0.09	0.15	0.39	0.92	1.00
SB^μ	0.10	0.18	0.48	0.97	1.00
$T(\hat{\rho}^\mu - 1)$	0.10	0.17	0.46	0.95	1.00
$\hat{\tau}^\mu$	0.08	0.13	0.34	0.87	1.00
$DK_S^\mu(\rho)$	0.12	0.20	0.53	0.97	1.00

The performance of the proposed tests is as follows. The test C_{1T} has similar power to C_1 in the non-stationary case and better power in the stationary case. The test C_A outperforms both of them. Furthermore, this test has a power close to the upper bound in the stationary case. These tests are all outperformed by B tests. These B tests have empirical power very close to the upper bound. Unreported computations show that the power of B tests comes mainly from the second derivatives of the Taylor expansion (statistics C_1 and C_{1T}), whereas the inclusion of the first derivative and the biased predictor helps to improve their performance. The application of the locally biased predictor to C_1 and C_{1T} increases their power in regions far from the unit circle. The addition of the first derivative, in the non-zero mean case, also with the locally biased predictor, increases the power in regions close to the unit circle.

The comparison with the remaining tests concludes that the tests with better power are

- Stationary case: B_S , P_T and $\hat{\tau}_{WS}$.
- Non-stationary case with non zero mean: B_N^μ , P_T^μ and $\hat{\tau}_{GLS}^\mu$.
- Non-stationary case with deterministic trend: B_N^τ , P_T^τ , $\hat{\tau}_{GLS}^\tau$ and $\hat{\tau}_{WS}^\tau$.

The performance of the competing tests in each of the three groups can be considered similar and close to the upper bound. The $\hat{\tau}_{GLS}$ test has different behavior in the stationary and non-stationary case. In the non-stationary case, the test has power close to $DK_N(\rho)$, but in the stationary case the performance is lower. This fact can be explained by the sub-optimal use of the information in the stationary case. It has previously been proved that, in the stationary case, an optimal test uses the information in both direct and reverse order, whereas $\hat{\tau}_{GLS}$ only uses the series in direct order. It seems reasonable that a modified $\hat{\tau}_{GLS}$ statistic that also uses the series in reverse order, as $\hat{\tau}_{WS}$ and B do, could attain a better performance. Likewise, $\hat{\tau}_{WS}$ has a better performance in the stationary case than in the non-stationary case. This fact could also be explained by a sub-optimal use of the information. It has been proved that, in the non-stationary case, an optimal test uses only the series in direct order, whereas $\hat{\tau}_{WS}$ tests uses both direct and reverse order. However, in the case of deterministic linear trend, it was proved that it is equivalent to use the series in direct or reverse order. This result can explain that $\hat{\tau}_{WS}^\tau$ has still good performance in the non-stationary case. The similar performance of B and P_T tests can be explained since both are alternative ways of building an optimal test under a local alternative.

6 The general ARMA case

This section introduces two procedures to extend the proposed tests to a general ARMA case. Both procedures incorporate the structure of u_t into the test statistics. The resulting test statistics have the same asymptotic distribution as in the AR(1) case. The first procedure summarizes the structure of u_t through a consistent estimator of κ^2 . The statistics are obtained by multiplying the estimator $\hat{\kappa}^2$ by each proposed statistic. These new family of statistics will be denoted as KC . The second procedure not only uses the estimator $\hat{\kappa}^2$ but also it incorporates the dynamics of u_t into the predictors. This second family of statistics will be denoted as PC .

Table 4: Empirical power at size 5%. $T = 100$. Stationary case. $d_t = \mu + \delta t$. 100,000 rep.

Test	ρ				
	0.97	0.95	0.90	0.80	0.70
C_1^τ	0.07	0.10	0.24	0.64	0.86
C_A^τ	0.07	0.10	0.25	0.69	0.91
B_S^τ	0.07	0.10	0.26	0.78	0.98
$\hat{\tau}_{WS}^\tau$	0.07	0.10	0.25	0.78	0.99
P_T^τ	0.07	0.10	0.26	0.77	0.98
$\hat{\tau}_{GLS}^\tau$	0.07	0.10	0.24	0.76	0.99
SB^τ	0.07	0.10	0.25	0.73	0.95
$T(\hat{\rho}^\tau - 1)$	0.07	0.09	0.23	0.74	0.98
$\hat{\tau}^\tau$	0.06	0.09	0.20	0.66	0.96
$DK_S^\tau(\rho)$	0.07	0.10	0.26	0.78	0.99

Table 5: Empirical power at size 5%. $T = 100$. Non-stationary case. $d_t = \mu$. 100,000 rep.

Test	ρ				
	0.97	0.95	0.90	0.80	0.70
C_1^μ	0.16	0.28	0.63	0.92	0.97
C_{1T}^μ	0.16	0.29	0.62	0.92	0.98
C_A^μ	0.15	0.25	0.57	0.97	1.00
B_N^μ	0.17	0.29	0.67	0.97	1.00
$\hat{\tau}_{WS}^\mu$	0.15	0.26	0.61	0.98	1.00
P_T^μ	0.17	0.29	0.66	0.96	0.99
$\hat{\tau}_{GLS}^\mu$	0.17	0.28	0.66	0.98	1.00
SB^μ	0.13	0.22	0.52	0.97	1.00
$T(\hat{\rho}^\mu - 1)$	0.12	0.19	0.47	0.95	1.00
$\hat{\tau}^\mu$	0.08	0.11	0.29	0.86	1.00
$DK_N^\mu(\rho)$	0.17	0.29	0.67	0.99	1.00

Table 6: Empirical power at size 5%. $T = 100$. Non-stationary case. $d_t = \mu + \delta t$. 100,000 rep.

Test	ρ				
	0.97	0.95	0.90	0.80	0.70
C_1^τ	0.07	0.11	0.29	0.72	0.91
C_A^τ	0.07	0.11	0.27	0.73	0.93
$B_N^\tau(c = 11)$	0.07	0.11	0.30	0.82	0.99
$\hat{\tau}_{WS}^\tau$	0.07	0.11	0.29	0.81	0.99
P_T^τ	0.07	0.11	0.30	0.82	0.98
$\hat{\tau}_{GLS}^\tau$	0.07	0.11	0.30	0.82	0.99
SB^τ	0.07	0.11	0.27	0.77	0.97
$T(\hat{\rho}^\tau - 1)$	0.07	0.10	0.24	0.74	0.99
$\hat{\tau}^\tau$	0.07	0.08	0.19	0.64	0.96
$DK_N^\tau(\rho)$	0.07	0.11	0.31	0.82	0.99

These two procedures have different justifications. The KC tests do not include more information than the minimum needed to use the same asymptotic results as in the $AR(1)$. Hence, these tests assume that their asymptotic properties are still applicable in finite samples. In particular, these tests assume that the unit root dominates the evolution of the series also in finite samples and, therefore, a test might not need much short-term information. Therefore, it would suffice to include a consistent estimation of κ^2 . The justification of the family of tests PC is apparent. If u_t has temporal dependency, the random walk predictor can be very inefficient even if the null hypothesis is true. A predictor based on the dynamics of u_t could improve the performance of the tests in finite samples.

6.1 KC tests

This section proposes modifications of the proposed statistics by using a consistent estimator of κ^2 . A new estimator of κ^2 , based on the efficient estimation of the ARMA model of u_t , will be considered. An estimator of κ^2 is

$$\hat{\kappa}^2 = \frac{\hat{\sigma}_u^2}{\hat{\omega}^2}, \quad (6.38)$$

where $\hat{\sigma}_u^2 = \sum_{t=2}^T (y_t - y_{t-1})^2 / T \xrightarrow{p} \sigma_u^2$. In the case of deterministic trend, the estimator $\hat{\delta}$ shown in (2.10) can be used. It can be noticed that $\hat{\sigma}_u^2$ are the denominators of C and $C(c)$. The family of statistics KC is easily derived. For instance, the generalization of C_1^μ is

$$KC_1^\mu = \hat{\kappa}^2 C_1^\mu = \frac{\sum_{t=1}^T (y_t - y_1)^2}{T^2 \hat{\omega}^2},$$

and the generalization of the LBI test is

$$KE^\mu = \frac{(y_T - y_1)^2}{T \hat{\omega}^2}.$$

Following the same fashion, the remaining statistics of previous sections can be generalized.

The KC tests need the consistent estimation of $\omega^2 = \sigma^2 \psi(1)^2$. One proposal for this estimator is due to Phillips (1987). It consists on the nonparametric estimator: $\hat{\omega}_c^2 = \sum_{m=-l_T}^{l_T} k\left(\frac{m}{l_T}\right) \hat{\gamma}_{\hat{a}}(m)$, where $\hat{\gamma}_x(m) = (T - m)^{-1} \sum_{t=|m|+1}^T (x_t - \bar{x})(x_{t-m} - \bar{x})$; with $k(\cdot)$ a kernel function and $\hat{a}_t$ the residuals of a regression of y_t with $(1, y_{t-1})$ or $(1, t, y_{t-1})$. The proper selection of the kernel and the number of lags l_T make this estimator consistent. With this approach, it is not necessary to know the dynamic structure of u_t . This property is useful when this structure is unknown or difficult to obtain. If u_t follows a known parametric model, or it can be properly identified, an estimator that uses such information would be preferred. Fuller (1976) proposes the following estimator for the $AR(p)$:

$$\hat{\omega}_{AR}^2 = \frac{\hat{\sigma}^2}{\left(1 - \sum_{j=1}^p \hat{\phi}_j^2\right)^2}, \quad (6.39)$$

where $\hat{\phi}_j$ are the least squares estimates of the error correction regression

$$\Delta y_t = \delta_0 + (\rho - 1)y_{t-1} + \sum_{j=1}^p \hat{\phi}_j^2 + a_t, \quad (6.40)$$

with $\delta_0 = \beta_0$ or $\delta_0 = \beta_0 + \beta_1 t$, depending on the deterministic component; and $\hat{\sigma}^2$ is the variance of the residuals. Said & Dickey (1984) show that this estimator is still consistent if u_t follows an ARMA process and p is properly chosen. Unit roots test statistics based on the estimator $\hat{\omega}_{AR}^2$ can have good properties if u_t follows a finite AR process or an ARMA process with moving average roots far from the unit circle. It is well known that unit roots tests based on this estimator can have high size distortion if the moving average component have positive roots of moderate or large size.

The estimator $\hat{\omega}_{AR}^2$ may be convenient for statistics based on the OLS estimator, like the Dickey-Fuller statistic $T(\hat{\rho} - 1)$, where $\hat{\rho}$ is estimated with (6.40). Since the new test statistics are not based on OLS, a better estimator of ω^2 can be proposed based on the efficient estimation of the model

$$\phi(B)(1 - \rho B)x_t = \theta(B)a_t,$$

where $x_t = y_t - \mu - \delta t$. The estimator $\hat{\omega}_{arma}^2$ will be

$$\hat{\omega}_{arma}^2 = \hat{\sigma}^2 \frac{(1 - \sum_{i=1}^q \hat{\theta}_i)^2}{(1 - \sum_{j=1}^p \hat{\phi}_j)^2}. \quad (6.41)$$

This estimator will be consistent if the estimators $\hat{\theta}_j$, $\hat{\phi}_i$, and the estimator of the residual variance are consistent.

6.2 Tests with efficient predictions. PC tests

This section uses the prediction-error interpretation to generalize the proposed tests to a general ARMA case. Let us see, first, the case with non-zero mean under the alternative. Let us assume that u_t admits the stationary and invertible ARMA representation $\phi(B)u_t = \theta(B)a_t$. Let us denote $\hat{\phi} = (\hat{\phi}_1, \dots, \hat{\phi}_p)'$ and $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_q)'$ to the parameter estimates. It will be shown that the comparison of the empirical PMSE and the expected one get to the same type of statistic tests than in the AR(1) case.

Similarly to the AR(1), an efficient predictor assuming the null hypothesis of unit root will be used. This implies a modellization in first differences. The forecasted value of y_t from y_{p+1} is, assuming known parameters, $\tilde{y}_{t|p+1} = y_{p+1} + \sum_{j=p+2}^t \Delta \tilde{y}_{j|p+1}$, with $\Delta \tilde{y}_{j|p+1} = \tilde{u}_{j|p+1} = E(u_j | u_{p+1})$. Since parameters are unknown it is necessary to use their estimates. The prediction of y_t from the first available observation will be $\hat{y}_{t|p+1} = y_{p+1} + \sum_{j=p+2}^t \Delta \hat{y}_{j|p+1}$, where $\Delta \hat{y}_{j|p+1} = \hat{u}_{j|p+1} = E(u_j | y_{p+1}, \hat{\phi}, \hat{\theta})$. The cumulated empirical prediction squared error will be

$$\sum_{t=p+2}^T e_{t|p+1}^2 = \sum_{t=p+2}^T (y_t - \hat{y}_{t|p+1})^2. \quad (6.42)$$

Since u_t is stationary, $\Delta y_j - \Delta \tilde{y}_{j|p+1} = u_j - \tilde{u}_{j|p+1} = a_j + \psi_1 a_{j-1} + \dots + \psi_{j-(p+2)} a_{p+2}$. Therefore, it can be obtained that $y_t - \hat{y}_{t|p+1} = a_t + (1 + \psi_1) a_{t-1} + \dots + (1 + \psi_1 + \dots + \psi_{t-(p+2)}) a_{p+2}$.

Therefore, the theoretical PMSE under the null hypothesis of a unit root is

$$E(y_t - \tilde{y}_{t|p+1})^2 = \sigma^2 \sum_{j=0}^{t-(p+2)} \left(\sum_{i=0}^j \psi_i \right)^2. \quad (6.43)$$

Since u_t is stationary, the coefficients ψ_j will decay exponentially with j . Therefore, as t increases, the term $\sum_{i=0}^{t-p-2} \psi_i$ approaches a constant value and the PMSE approaches the value $\omega^2(t-p-2)$, where $\omega^2 = \sigma^2\psi(1)^2$. Hence, the cumulated PMSE is, approximately,

$$\sum_{t=p+2}^T E(y_t - \tilde{y}_{t|p+1})^2 \approx \omega^2 \{1 + 2 + \dots + (T-p-1)\} = \omega^2 \frac{(T-p)(T-p-1)}{2}. \quad (6.44)$$

The parameter ω^2 is unknown and an estimate should be used. The ratio between the cumulated empirical PMSE (6.42) and the theoretical expected values (6.44) suggests the statistic PC_1^μ :

$$PC_1^\mu = \frac{\sum_{t=p+2}^T (y_t - \hat{y}_{t|p+1})^2}{(T-p)^2 \hat{\omega}^2}. \quad (6.45)$$

This statistic is similar to KC_1^μ , except that the value y_1 is replaced by the prediction $\hat{y}_{t|p+1}$. Also it is divided by $(T-p)^2$ instead of T^2 since p observations are lost. In the deterministic trend case it is verified that $\Delta y_t = \delta + u_t$. The parameter δ is unknown but the estimator $\hat{\delta}$ proposed in (2.10) can be used. Also, and asymptotically equivalent, it could be jointly estimated with $\hat{\phi}$ and $\hat{\theta}$. Let us denote $\hat{y}_{t|p+1}^\tau$ to the predictions in this deterministic trend case. The test statistic PC_1^τ is

$$PC_1^\tau = \frac{\sum_{t=p+2}^T (y_t - \hat{y}_{t|p+1}^\tau)^2}{(T-p)^2 \hat{\omega}^2}. \quad (6.46)$$

Similarly, the corresponding expressions for the remaining statistics can be obtained. In the B tests, series can previously be detrended by GLS using the matrix $\mathbf{\Omega}_S^{-1}(\rho_c)$ or $\mathbf{\Omega}_N^{-1}(\rho_c)$ and the predictor can be built using the detrended series. For instance, in the ARMA(1,1) case, and considering $a_t = 0$ for $t < 1$, the prediction from the first observation in the stationary case, $\hat{y}_{t|1-S}^\mu$, will be $\hat{y}_{t|1-S}^\mu = \rho_c^{t-1}(\rho_c - \hat{\theta})\hat{y}_{1|S}^\mu$, with $\hat{\theta}$ a root-T consistent estimator of θ . The following theorem state that the family of statistics PC have the same limiting distributions of KC . Proof is in the appendix.

Theorem 2 *Let y_t be the process (1.2) with $\rho = 1$. Let u_t be the stationary and invertible ARMA(p, q). Let $\hat{\phi} = (\hat{\phi}_1, \dots, \hat{\phi}_p)'$ and $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_q)'$ be estimators of $\phi = (\phi_1, \dots, \phi_p)'$ and $\theta = (\theta_1, \dots, \theta_q)'$, respectively, such that $\hat{\phi} - \phi = O_p(T^{-1/2})$ and $\hat{\theta} - \theta = O_p(T^{-1/2})$. Then,*

$$KC - PC \xrightarrow{d} 0. \quad (6.47)$$

6.3 Finite sample performance. ARMA(1,1) case.

This section presents a Monte Carlo experiment to evaluate the finite sample performance of the proposed tests. The model used in the experiment is

$$y_t = \rho y_{t-1} + a_t - \theta a_{t-1}, \quad (6.48)$$

with $a_t \sim N(0,1)$ and sample sizes $T = 50, 100$. Only $T = 100$ is reported. Conclusions are similar in $T = 50$. Only the better tests found in section 5 are considered. Experiments have been made for both the non-stationary case ($y_1 = a_1$) and the stationary case. The first observation in the stationary case ($\rho < 1$) has variance equal to $(1 + \theta^2 - 2\theta\rho)/(1 - \rho^2)$. Besides, 50 initial observations have been dropped to eliminate initial transients.

To better simulate the behavior of the proposed tests and also the estimator $\hat{\omega}_{arma}^2$, an analyst that searches for the best model in each replication would be necessary. Instead, it will be assumed that an ARMA(1,1) is estimated to evaluate $\hat{\omega}_{arma}^2$. The estimation method is nonlinear least squares (Gauss-Newton) with series previously detrended by OLS. This approach has proven to offer much better results than the joint estimation of the deterministic components and the ARMA parameters. The estimates have been searched around an initial set of values. Initial values have been obtained using the relation between these ARMA parameters and the π -weights of the autoregressive representation. This representation holds $\pi(B)x_t = a_t$, $\pi(B) = (1 - \rho B)(1 - \theta B)^{-1}$, and, therefore, $\pi_1 = \rho - \theta$ and $\pi_2 = \theta(\rho - \theta)$. It is verified that $\theta = \pi_2/\pi_1$ and $\rho = \pi_1 - \theta$. To apply this result, an AR(6) has been estimated by OLS to the detrended series. It has also been investigated, as initial values, the corresponding to the null hypotheses, but the performance of the tests was worst. The number of replications that were discarded due to lack of convergence was very small. Different estimators of the moving average parameter θ in the numerator the B tests have been investigated. Namely, the exact maximum likelihood of the IMA(1,1) model, the exact maximum likelihood of the ARMA(1,1) model restricted to $\rho = \rho_c$, and the nonlinear least squares of the unrestricted ARMA(1,1) (the same as in $\hat{\omega}_{arma}^2$). The performance of the tests was very similar with any of the estimation methods. Only results from the unrestricted ARMA(1,1) estimator are reported.

The P_T tests have been studied with both the new estimator $\hat{\omega}_{arma}^2$ and also with $\hat{\omega}_{AR}^2$, previously proposed by their authors Elliott *et al.* (1996) and Elliott (1997). To evaluate $\hat{\omega}_{AR}^2$, an augmented error correction model using the original series y_t has been used. The number of lags has been selected with BIC criteria and two different restrictions. First, BIC criteria has been constrained to choose values in the range $0 \leq p \leq 8$ to avoid large values of the autoregressive order p . It is well known that, for large and positive values of θ , BIC criteria tends to be small. To avoid this effect, a second experiment has been performed with BIC restricted in the range $3 \leq p \leq 8$. Only results of this second restriction are reported. The restriction $0 \leq p \leq 8$ produces size distortions than, in some cases, can overpass 0.60. Similarly, the BIC criteria with the same restrictions on p has been used in $\hat{\tau}_{WS}$ and $\hat{\tau}_{GLS}$.

Tables 7 and 8 contains the proportion of times a unit root has been detected for different values of ρ and θ in 5,000 replications. The column AR(1) shows the results for the AR(1) case reported in tables 3 and 4. The comparison of this column with the remaining columns shows the distortion due to the dynamic of u_t . A main feature is that the proposed tests PB , KB , and P_T when used with the proposed estimator $\hat{\omega}_{arma}^2$ have similar behavior both in size and power. These three tests have very small size distortion. The empirical size is very close to the nominal level except for very high positive values of θ . Even in this extreme case, the size distortion is much lower than the competing tests based in autoregressive approximations. The tests with smaller size distortion is PB . Therefore, improving the efficiency of the predictors can help to avoid size

distortion. Regarding only tests that rely on autoregressive approximations (P_T with $\hat{\omega}_{AR}^2$, $\hat{\tau}_{WS}$, and $\hat{\tau}_{GLS}$), the test that suffers greater size distortion with $\theta = 0.8$ is $\hat{\tau}_{WS}$, but it is also, together with $\hat{\tau}_{GLS}$ in the non-stationary case, the test with smallest size distortion in the remaining cases.

In order to compare the power performance, it should be distinguished between $\theta \leq 0$ and $\theta > 0$. For $\theta \leq 0$, the proposed tests and also P_T with $\hat{\omega}_{arma}^2$ show a very high power even for values of ρ far from the unit circle. The comparison of their empirical power with the AR(1) case shows a lower distortion than the remaining tests. These remaining tests only overpass these three tests in regions close to the unit circle and only if their size distortion is bigger. For $\theta > 0$, the proposed tests and P_T with $\hat{\omega}_{arma}^2$ show a deterioration in power with respect to the AR(1) case. This effect can be explained by the inefficient estimation of the ARMA(1,1) parameters when ρ is not much different to θ . The PB tests have slightly better power than KB and P_T . Hence, the inclusion of an efficient predictor can also improve the power of the tests, but the gain may be moderate. The remaining tests have a power much greater than the AR(1) case. This high power is due to their size distortion. Therefore, as the process approaches white noise, all tests deteriorate their performance. The proposed tests, due to the high variance of the estimations, lose their high power. The remaining tests, due to their inability to model the MA component, experience a very high size distortion.

If the analyst has chosen the null hypothesis of unit root, the better option would be the proposed tests, since they assure the size much better than their competitors and still have power. Otherwise, if the analyst is willing to admit size distortions as large as 0.40, he would be well advised to choose a test with stationarity as a null hypothesis.

7 Concluding remarks

This paper analyzes the relationship between the properties of the PMSE and the optimal detection of a unit root and helps to understand how optimal tests use the information. According to this relationship, tests that are most powerful in a neighborhood of the locality are proposed. It is shown that these MPNL tests are based on the forecasting performance of a predictor that assumes a unit root. If the initial value is extracted from its conditional distribution, the predictions are made from the first observation. Therefore, the evolution of the prediction errors of each observation from the first value contains valuable information about the presence of a unit root. These prediction errors are implicitly used by some others tests, as tests based on maximum likelihood estimation (MLE) or GLS under a fixed local alternative (see Remark 1.1 in Pantula *et al.*, 1994). This can explain why these tests, which are not explicitly optimal, overpass those based on OLS. If the initial value is such that the process is stationary under the alternative, MPNL tests use the information of both the original series and its time reversal. Hence, the prediction errors are evaluated from both extreme values. Following the same arguments than in Pantula *et al.* and the proof of proposition 3.2, it can be proved that these prediction errors are also implicitly used by tests based on unconditional MLE or GLS under a fixed local alternative. Time reversed series are explicitly used by tests based on the simple and weighted symmetric estimators. Again, the more optimal use of the information can explain that these tests overpass those based on OLS.

The prediction-error interpretation is used to modify the proposed MPNL tests with a locally biased predictor. The power of the resulting B tests is similar to the Gaussian power envelope. In the general ARMA case, the proposed tests admit generalizations based on the best model instead of autoregressive approximations. This best model can be used both to estimate ω^2 and to improve the efficiency of the predictors. Computations show that $\hat{\omega}_{arma}^2$ improves the performance of the tests when compared with the competitors. In the analyzed ARMA(1,1) case, the use of a more efficient predictor in PB tests only produces a moderate improvement. Nevertheless, the use of a more efficient predictor is recommended, since it may diminishes the risk of low performance in more complex models than reported here.

The tests proposed in this article have both intuitive interpretation and optimal properties. Besides, they have excellent performance both in power and in size. These properties make them a convenient tool for unit root detection.

Table 7: Empirical power at size 5%. $T = 100$. Initial conditions: non-stationary. 5,000 rep.

Test	ρ	AR(1)	θ				
			-0.8	-0.5	0.0	0.5	0.8
PB_N^μ	1.00	0.05	0.05	0.06	0.07	0.05	0.11
	0.95	0.29	0.34	0.34	0.33	0.19	0.31
	0.90	0.67	0.75	0.74	0.67	0.36	0.52
	0.80	0.97	1.00	0.99	0.94	0.55	0.70
KB_N^μ	1.00	0.05	0.06	0.06	0.07	0.05	0.11
	0.95	0.29	0.34	0.35	0.33	0.19	0.30
	0.90	0.67	0.75	0.74	0.67	0.35	0.51
	0.80	0.97	1.00	0.99	0.94	0.54	0.70
$P_T^\mu(\hat{\omega}_{arma}^2)$	1.00	0.05	0.05	0.06	0.07	0.05	0.11
	0.95	0.29	0.33	0.34	0.33	0.19	0.30
	0.90	0.66	0.75	0.74	0.66	0.35	0.51
	0.80	0.96	1.00	0.99	0.92	0.52	0.69
$P_T^\mu(\hat{\omega}_{AR}^2)$	1.00	0.05	0.14	0.09	0.09	0.09	0.38
	0.95	0.29	0.49	0.45	0.39	0.36	0.75
	0.90	0.66	0.75	0.74	0.68	0.62	0.88
	0.80	0.96	0.92	0.95	0.90	0.85	0.97
$\hat{\tau}_{WS}^\mu$	1.00	0.05	0.08	0.06	0.06	0.09	0.52
	0.95	0.26	0.31	0.28	0.24	0.34	0.92
	0.90	0.61	0.56	0.53	0.49	0.64	0.99
	0.80	0.98	0.82	0.87	0.85	0.96	1.00
$\hat{\tau}_{GLS}^\mu$	1.00	0.05	0.10	0.07	0.07	0.10	0.47
	0.95	0.28	0.39	0.35	0.31	0.39	0.83
	0.90	0.66	0.66	0.65	0.59	0.67	0.91
	0.80	0.98	0.89	0.92	0.89	0.90	0.95
PB_N^τ	1.00	0.05	0.07	0.07	0.07	0.06	0.26
	0.95	0.11	0.14	0.14	0.15	0.10	0.38
	0.90	0.30	0.36	0.37	0.33	0.18	0.54
	0.80	0.82	0.88	0.88	0.76	0.40	0.62
KB_N^τ	1.00	0.05	0.07	0.07	0.07	0.06	0.26
	0.95	0.11	0.15	0.15	0.15	0.10	0.38
	0.90	0.30	0.37	0.38	0.33	0.18	0.54
	0.80	0.82	0.88	0.88	0.76	0.39	0.62
$P_T^\tau(\hat{\omega}_{arma}^2)$	1.00	0.05	0.07	0.07	0.07	0.06	0.26
	0.95	0.11	0.14	0.15	0.15	0.10	0.38
	0.90	0.30	0.37	0.37	0.33	0.18	0.54
	0.80	0.82	0.88	0.88	0.76	0.40	0.62
$P_T^\tau(\hat{\omega}_{AR}^2)$	1.00	0.05	0.26	0.18	0.14	0.14	0.53
	0.95	0.11	0.41	0.32	0.27	0.25	0.70
	0.90	0.30	0.60	0.53	0.47	0.42	0.84
	0.80	0.82	0.81	0.82	0.77	0.75	0.97
$\hat{\tau}_{WS}^\tau$	1.00	0.05	0.09	0.06	0.05	0.09	0.65
	0.95	0.11	0.18	0.14	0.11	0.18	0.83
	0.90	0.29	0.32	0.26	0.23	0.36	0.96
	0.80	0.81	0.59	0.57	0.54	0.78	1.00
$\hat{\tau}_{GLS}^\tau$	1.00	0.05	0.08	0.05	0.04	0.07	0.46
	0.95	0.11	0.15	0.12	0.09	0.13	0.62
	0.90	0.30	0.29	0.23	0.20	0.28	0.77
	0.80	0.82	0.56	0.53	0.50	0.63	0.88

Table 8: Empirical power at size 5%. $T = 100$. Initial conditions: stationary. 5,000 rep.

Test	ρ	AR(1)	θ				
			-0.8	-0.5	0.0	0.5	0.8
PB_S^μ	1.00	0.05	0.06	0.07	0.06	0.05	0.11
	0.95	0.19	0.24	0.25	0.25	0.16	0.29
	0.90	0.52	0.60	0.60	0.56	0.33	0.52
	0.80	0.97	0.97	0.97	0.93	0.55	0.71
KB_S^μ	1.00	0.05	0.07	0.07	0.06	0.05	0.11
	0.95	0.19	0.24	0.25	0.25	0.15	0.29
	0.90	0.52	0.61	0.60	0.56	0.32	0.52
	0.80	0.97	0.98	0.97	0.92	0.54	0.71
$P_T^\mu(\hat{\omega}_{arma}^2)$	1.00	0.05	0.07	0.07	0.06	0.04	0.11
	0.95	0.20	0.23	0.25	0.24	0.14	0.29
	0.90	0.52	0.60	0.60	0.55	0.31	0.52
	0.80	0.97	0.98	0.97	0.93	0.54	0.71
$P_T^\mu(\hat{\omega}_{AR}^2)$	1.00	0.05	0.14	0.12	0.09	0.09	0.43
	0.95	0.20	0.37	0.33	0.29	0.29	0.77
	0.90	0.52	0.63	0.63	0.56	0.59	0.95
	0.80	0.97	0.86	0.90	0.89	0.91	1.00
$\hat{\tau}_{WS}^\mu$	1.00	0.05	0.07	0.07	0.05	0.08	0.53
	0.95	0.20	0.24	0.21	0.18	0.28	0.89
	0.90	0.52	0.48	0.46	0.42	0.61	0.99
	0.80	0.97	0.78	0.82	0.82	0.95	1.00
PB_S^τ	1.00	0.05	0.05	0.06	0.06	0.05	0.26
	0.95	0.10	0.11	0.12	0.12	0.09	0.37
	0.90	0.26	0.26	0.28	0.26	0.18	0.55
	0.80	0.78	0.77	0.78	0.69	0.39	0.63
KB_S^τ	1.00	0.05	0.07	0.07	0.06	0.05	0.26
	0.95	0.10	0.13	0.13	0.12	0.09	0.37
	0.90	0.26	0.29	0.30	0.27	0.18	0.54
	0.80	0.78	0.81	0.80	0.69	0.39	0.63
$P_T^\tau(\hat{\omega}_{arma}^2)$	1.00	0.05	0.07	0.07	0.07	0.05	0.27
	0.95	0.10	0.13	0.14	0.13	0.09	0.37
	0.90	0.26	0.32	0.33	0.29	0.18	0.55
	0.80	0.77	0.82	0.83	0.72	0.39	0.63
$P_T^\tau(\hat{\omega}_{AR}^2)$	1.00	0.05	0.26	0.20	0.15	0.14	0.55
	0.95	0.10	0.36	0.30	0.24	0.22	0.67
	0.90	0.26	0.55	0.49	0.42	0.42	0.85
	0.80	0.77	0.79	0.77	0.73	0.75	0.97
$\hat{\tau}_{WS}^\tau$	1.00	0.05	0.10	0.07	0.06	0.10	0.64
	0.95	0.10	0.16	0.12	0.10	0.17	0.81
	0.90	0.25	0.29	0.24	0.20	0.35	0.95
	0.80	0.78	0.57	0.54	0.51	0.77	1.00

APPENDIX

Proof of theorem 1:

The asymptotic distribution of C_1^μ is a straightforward application of the functional limit theorem (FCLT) to $S_t = \sum_{j=2}^t u_j$. To obtain the asymptotic distribution of C_1^τ , it is convenient to rewrite the statistic introducing the term $(\hat{\delta} - \delta)$. Applying that

$$\hat{\delta} = \frac{y_T - y_1}{T - 1} = \delta + \frac{\sum_{t=2}^T u_j}{T - 1}$$

it can be obtained that $\sqrt{T}(\hat{\delta} - \delta) \xrightarrow{d} \sigma\psi(1)W(1)$. Applying this result and making use of the FCLT and that

$$\frac{\sum_{t=1}^T t^k}{T^{(k+1)}} \longrightarrow \frac{1}{k+1}, \quad k = 0, 1, \dots$$

as $T \rightarrow \infty$, the expression (2.16) can be obtained. To obtain the asymptotic distribution of C_{1T}^μ , the following decomposition can be used:

$$\sum_{t=1}^{T-1} \{(y_t - y_1) - (y_T - y_1)\}^2 = \sum_{t=1}^{T-1} (y_t - y_1)^2 + (T-1)(y_T - y_1)^2 - 2(y_T - y_1) \sum_{t=2}^T (y_t - y_1),$$

and making use of the FCLT the result holds. To obtain the asymptotic distribution of C_A^μ it can be defined $Z(r, \lambda)$, $\lambda < r$, as

$$Z(r, \lambda) = \frac{1}{T} \frac{\sum_{[T\lambda]}^{[Tr]} u_t}{\sqrt{T - (r - \lambda)}} \quad (1)$$

where $[\cdot]$ denotes the integer part of the argument. The partial sum $\sum_{[T\lambda]}^{[Tr]} u_t$ is the prediction error of $y_{[Tr]}$ from $y_{[T\lambda]}$ using a random walk. The term $T - (r - \lambda)$ is the number of partial sums with the same number of terms in the sample. For instance, there are $T - 1$ partial sums with only one term, $T - 2$ partial sums with two terms and so on. Applying FCLT it holds that

$$\begin{aligned} \int_0^1 \int_0^r Z^2(r, \lambda) d\lambda dr &= T^{-2} \sum_{t=1}^{T-1} \sum_{s=t+1}^T \{T - (s - t)\}^{-1} (y_s - y_t)^2 \\ &\xrightarrow{d} \sigma^2 \psi^2(1) \int_0^1 \int_0^r (1 - (r - \lambda))^{-1} W^2(r - \lambda) dr d\lambda. \end{aligned} \quad (2)$$

Using that $\sum_{t=1}^{T-1} \sum_{s=t+1}^T (s - t)^2 \{T - (s - t)\}^{-1} = \sum_{t=1}^{T-1} \sum_{s=t+1}^T (s - t) = \sum_{t=1}^{T-1} t^2$ and applying the same arguments than in C_1^τ , the expression (2.19) is obtained. $\square$

Proof of proposition 4.1

Only the proof for the statistic B_N^τ is shown. The proof for the remaining statistics uses the same arguments. It is necessary to analyze the convergence of $T^{-2} \sum_{t=2}^T (y_{t|N}^\tau - \rho_c^{t-1} y_{1|N}^\tau)^2$. The following decomposition can be used:

$$\begin{aligned} \sum_{t=2}^T (y_{t|N}^\tau - \rho_c^{t-1} y_{1|N}^\tau)^2 &= \sum_{t=2}^T (y_{t|N}^\tau - y_{1|N}^\tau)^2 + \sum_{t=2}^T (1 - \rho_c^{t-1})^2 y_{1|N}^\tau^2 \\ &\quad + 2 \sum_{t=2}^T (y_{t|N}^\tau - y_{1|N}^\tau) (1 - \rho_c^{t-1}) y_{1|N}^\tau. \end{aligned} \quad (3)$$

Since $y_{1|N}^\tau = O_p(1)$, the second term of (3) converges to zero in probability when divided by T^2 . It can be obtained, also, that

$$(y_{t|N}^\tau - y_{1|N}^\tau) = y_t - y_1 - \delta(t-1) + (\delta - \hat{\delta})(t-1) + (\hat{\delta} - \hat{\delta}_N)(t-1).$$

Since $(\hat{\delta} - \hat{\delta}_N) = O_p(T^{-1})$, $(\delta - \hat{\delta}) = O_p(T^{-1/2})$ and $\sum y_t - y_1 - \delta(t-1) = O_p(t^{3/2})$ it can be seen that

$$T^{-2} \sum_{t=2}^T (y_{t|N}^\tau - y_{1|N}^\tau) \xrightarrow{p} 0$$

and, then, the third term in (3) also vanishes asymptotically when divided by T^2 . Using these arguments, the first term in (3) verifies

$$\frac{\sum_{t=2}^T (y_{t|N}^\tau - \rho_c^{t-1} y_{1|N}^\tau)^2}{T^2} - \frac{\sum_{t=2}^T \{y_t - y_1 - \delta(t-1)\}}{T^2} \xrightarrow{d} 0$$

and the proposition holds. $\square$

Proof of Proposition 3.1:

The GLS residuals of the numerator of S_N are $\hat{x}_{N,t} = y_t - \hat{\mu}_N$. For simplicity of notation, the argument may be omitted when it is not necessary. Then,

$$\hat{\mu}_N = \frac{(1 + \rho^2 - \rho)y_1 + (1 - \rho)y_T + \sum_{t=2}^{T-1} (1 - \rho)^2 y_t}{1 + (T-1)(1 - \rho)^2}. \quad (4)$$

Likewise, the residuals of the denominator of S_N are $\hat{x}_{1,t} = y_t - \hat{\mu}_1$, where $\hat{\mu}_1$ is the GLS estimator with the matrix $\mathbf{\Omega}_1$. Then, $\hat{\mu}_1 = y_1$. The first derivative is

$$S'_N(1) = \frac{\frac{\partial}{\partial \rho} \hat{\mathbf{x}}'_N \mathbf{\Omega}_N^{-1} \hat{\mathbf{x}}_N \Big|_{\rho=1}}{\hat{\mathbf{x}}'_1 \mathbf{\Omega}_1^{-1} \hat{\mathbf{x}}_1},$$

where it can be verified that $\hat{\mathbf{x}}'_1 \mathbf{\Omega}_1^{-1} \hat{\mathbf{x}}_1 = \sum_{t=2}^T (y_t - y_{t-1})^2$. Then, the numerator of $S'_N(1)$ can be written as

$$\frac{\partial}{\partial \rho} \hat{\mathbf{x}}'_N \mathbf{\Omega}_N^{-1} \hat{\mathbf{x}}_N \Big|_{\rho=1} = 2 \frac{\partial \hat{\mathbf{x}}'_N}{\partial \rho} \mathbf{\Omega}_N^{-1} \hat{\mathbf{x}}_N \Big|_{\rho=1} + \hat{\mathbf{x}}'_N \frac{\partial \mathbf{\Omega}_N^{-1}}{\partial \rho} \hat{\mathbf{x}}_N \Big|_{\rho=1}.$$

Since

$$\frac{\partial \hat{\mathbf{x}}'_N}{\partial \rho} = -\frac{\partial \hat{\mu}_N}{\partial \rho} \mathbf{e}'$$

and also

$$\mathbf{e}' \mathbf{\Omega}_N^{-1} \hat{\mathbf{x}}'_N \Big|_{\rho=1} = 0, \quad (5)$$

it can be obtained that

$$\frac{\partial}{\partial \rho} \hat{\mathbf{x}}'_N \mathbf{\Omega}_N^{-1} \hat{\mathbf{x}}_N \Big|_{\rho=1} = \hat{\mathbf{x}}'_N \frac{\partial \mathbf{\Omega}_N^{-1}}{\partial \rho} \hat{\mathbf{x}}_N \Big|_{\rho=1}.$$

From (3.20) it can be obtained that

$$\frac{\partial \mathbf{\Omega}_N^{-1}}{\partial \rho} \Big|_{\rho=1} = 2\mathbf{L}'\mathbf{L} - (\mathbf{L} + \mathbf{L}') = (\mathbf{I} - \mathbf{L})'(\mathbf{I} - \mathbf{L}) - \text{diag}(0, \dots, 0, 1) \quad (6)$$

where diag expresses a diagonal matrix. Also from (4) it holds that $\hat{\mathbf{x}}_{N,t}(1) = (y_t - y_1)$. Therefore, it can be obtained that

$$S'_N(1) = 1 - \frac{(y_T - y_1)^2}{\sum_{t=2}^T (y_t - y_{t-1})^2}.$$

For the second derivative, it can be written that

$$\begin{aligned} \frac{\partial^2}{\partial \rho^2} \hat{\mathbf{x}}'_N \boldsymbol{\Omega}_N^{-1} \hat{\mathbf{x}}_N &= 2 \frac{\partial^2 \hat{\mathbf{x}}'_N}{\partial \rho^2} \boldsymbol{\Omega}_N^{-1} \hat{\mathbf{x}}_N + 4 \frac{\partial \hat{\mathbf{x}}'_N}{\partial \rho} \frac{\partial \boldsymbol{\Omega}_N^{-1}}{\partial \rho} \hat{\mathbf{x}}_N \\ &\quad + 2 \frac{\partial \hat{\mathbf{x}}'_N}{\partial \rho} \boldsymbol{\Omega}_N^{-1} \frac{\partial \hat{\mathbf{x}}_N}{\partial \rho} + \hat{\mathbf{x}}'_N \frac{\partial^2 \boldsymbol{\Omega}_N^{-1}}{\partial \rho^2} \hat{\mathbf{x}}_N. \end{aligned} \quad (7)$$

Applying (5), the first term in (7) is null at $\rho = 1$. Also, by (6) and applying that

$$\left. \frac{\partial \hat{\mu}_N}{\partial \rho} \right|_{\rho=1} = (y_1 - y_T),$$

it can be obtained that

$$4 \left. \frac{\partial \hat{\mathbf{x}}'_N}{\partial \rho} \frac{\partial \boldsymbol{\Omega}_N^{-1}}{\partial \rho} \hat{\mathbf{x}}_N \right|_{\rho=1} = -4(y_T - y_1)^2.$$

Similarly, since $\mathbf{e}' \boldsymbol{\Omega}_N^{-1}(1) \mathbf{e} = 1$ it holds that

$$2 \left. \frac{\partial \hat{\mathbf{x}}'_N}{\partial \rho} \boldsymbol{\Omega}_N^{-1} \frac{\partial \hat{\mathbf{x}}_N}{\partial \rho} \right|_{\rho=1} = 2(y_T - y_1)^2.$$

To solve the last term in (7) it can be applied that

$$\left. \frac{\partial^2 \boldsymbol{\Omega}_N^{-1}}{\partial \rho^2} \right|_{\rho=1} = 2 \mathbf{L}' \mathbf{L}, \quad (8)$$

and, therefore,

$$\left. \hat{\mathbf{x}}'_N \frac{\partial^2 \boldsymbol{\Omega}_N^{-1}}{\partial \rho^2} \hat{\mathbf{x}}_N \right|_{\rho=1} = 2 \sum_{t=2}^{T-1} (y_t - y_1)^2$$

and proposition holds. $\square$

Proof of Proposition 3.2:

The GLS residuals of the numerator of S_S are $\hat{x}_{S,t}(\rho) = y_t - \hat{\mu}_S(\rho)$. Then (ommiting the argument),

$$\hat{\mu}_S = \frac{(1 - \rho)y_1 + (1 - \rho)y_T + \sum_{t=2}^{T-1} (1 - \rho)^2 y_t}{2(1 - \rho) + (T - 2)(1 - \rho)^2} \quad (9)$$

$$= \frac{y_1 + y_T + \sum_{t=2}^{T-1} (1 - \rho)y_t}{2 + (T - 2)(1 - \rho)} \quad (10)$$

$$(11)$$

The denominator of S_S is the same than in the previous case. Following the same arguments than in proposition 3.1 it holds that

$$\mathbf{e}'\boldsymbol{\Omega}_S^{-1}\hat{\mathbf{x}}'_S\Big|_{\rho=1} = 0. \quad (12)$$

Also, by (3.22), it holds that

$$\frac{\partial \boldsymbol{\Omega}_S^{-1}}{\partial \rho}\Big|_{\rho=1} = 2\mathbf{L}'\mathbf{L} - (\mathbf{L} + \mathbf{L}') - 2\mathbf{e}\mathbf{e}' = (\mathbf{I} - \mathbf{e}\mathbf{e}' - \mathbf{L})'(\mathbf{I} - \mathbf{e}\mathbf{e}' - \mathbf{L}) - \text{diag}(1, 0, \dots, 0, 1). \quad (13)$$

From (9) it holds that $\hat{x}_{S,t}(1) = y_t - (y_1 + y_T)/2$. Then, it can be obtained that

$$S'_S(1) = 1 - \frac{(y_T - y_1)^2}{2 \sum_{t=2}^T (y_t - y_{t-1})^2}.$$

Likewise, in the second derivative, it can be obtained that

$$\begin{aligned} \frac{\partial^2}{\partial \rho^2} \hat{\mathbf{x}}'_S \boldsymbol{\Omega}_S^{-1} \hat{\mathbf{x}}_S &= 2\partial^2 \hat{\mathbf{x}}'_S \partial \rho^2 S'_S \boldsymbol{\Omega}_S^{-1} \hat{\mathbf{x}}_S + 4 \frac{\partial \hat{\mathbf{x}}'_S}{\partial \rho} \frac{\partial \boldsymbol{\Omega}_S^{-1}}{\partial \rho} \hat{\mathbf{x}}_S \\ &\quad + 2 \frac{\partial \hat{\mathbf{x}}'_S}{\partial \rho} \boldsymbol{\Omega}_S^{-1} \frac{\partial \hat{\mathbf{x}}_S}{\partial \rho} + \hat{\mathbf{x}}'_S \frac{\partial^2 \boldsymbol{\Omega}_S^{-1}}{\partial \rho^2} \hat{\mathbf{x}}_S. \end{aligned} \quad (14)$$

Similarly to the non-stationary case, it can be obtained that the first term in (14) is null at $\rho = 1$. Applying (13) it can also be seen that the second term in (14) is also null in $\rho = 1$. Since $\mathbf{e}'\boldsymbol{\Omega}_S^{-1}(1)\mathbf{e} = 0$ the third term is also null at unity. To solve the fourth term of (14) it can be verified that

$$\frac{\partial^2 \boldsymbol{\Omega}_S^{-1}}{\partial \rho^2}\Big|_{\rho=1} = 2\mathbf{L}'\mathbf{L} - 2\mathbf{e}\mathbf{e}' = \text{diag}(0, 2, 2, \dots, 2, 0), \quad (15)$$

and, then,

$$\hat{\mathbf{x}}'_S \frac{\partial^2 \boldsymbol{\Omega}_S^{-1}}{\partial \rho^2} \hat{\mathbf{x}}_S \Big|_{\rho=1} = \sum_{t=2}^{T-1} (y_t - y_1)^2 + \sum_{t=2}^{T-1} (y_T - y_t)^2 - 2 \sum_{t=2}^{T-1} (y_t - y_1)(y_T - y_t),$$

and proposition holds. $\square$

Proof of Proposition 3.3:

The proof will be developed jointly for the stationary and non-stationary case and there will be notational differences only when it is necessary. The GLS residuals in the numerator of S_N and S_S are $\hat{x}_t = y_t - \hat{\mu} - \hat{\delta}t$, where $\hat{\mu}$ and $\hat{\delta}$ are the GLS estimator with the corresponding matrix. Likewise, the residuals of the denominator are $\hat{x}_{1,t} = y_t - \hat{\mu}_1 - \hat{\delta}t$, where $\hat{\mu}_1$ and $\hat{\delta}$ are the GLS estimator with the matrix $\boldsymbol{\Omega}_1$. Then, $\hat{\mu}_1 = y_1 - \delta$ and $\hat{\delta} = (y_T - y_1)(T-1)^{-1}$, both in the stationary and non-stationary case. It can be verified that $\hat{\mathbf{x}}'_1 \boldsymbol{\Omega}_1^{-1} \hat{\mathbf{x}}_1 = \sum_{t=2}^T (y_t - y_{t-1} - \hat{\delta})^2$. The numerator of $S'(1)$ can be written as

$$\frac{\partial}{\partial \rho} \hat{\mathbf{x}}' \boldsymbol{\Omega}^{-1} \hat{\mathbf{x}} \Big|_{\rho=1} = 2 \frac{\partial \hat{\mathbf{x}}'}{\partial \rho} \boldsymbol{\Omega}^{-1} \hat{\mathbf{x}} \Big|_{\rho=1} + \hat{\mathbf{x}}' \frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \rho} \hat{\mathbf{x}} \Big|_{\rho=1}. \quad (16)$$

It holds that the first derivative of $\hat{\delta}$ evaluated at $\rho = 1$ is null both in the stationary and non-stationary case. It can also be checked that the first derivative of $\hat{\mu}$ is also null at $\rho = 1$, but only in the non-stationary case. It can, then, be verified that the first derivative of $\hat{\mathbf{x}}'$ at $\rho = 1$ is a vector with all elements equal to zero. Therefore, the first term at the right side of (16) is null. Similarly, applying (6) and (13) it can be seen that the second term of (16) is equal to one. For the second derivative, there is a similar decomposition to 7). Following the same arguments than in the first derivative it can be seen that the second and third terms of this decomposition are zero. The term $(\partial^2 \hat{\mathbf{x}}' / \partial \rho^2) \boldsymbol{\Omega}^{-1} \Big|_{\rho=1}$ is a vector with all elements equal to zero except the first and the last one. Also the term $\hat{\mathbf{x}}(1)$ is a vector with first and last elements equal to zero. Therefore, the first term of (16) is also null. By (8) and (15) we obtain that the forth term is $2 \sum_{t=2}^{T-1} \left\{ y_t - y_1 - \delta(t-1) \right\}^2$, and the propositions hold. $\square$

Proof of theorem 2:

The proof is developed for the statistic PC_1^μ . The remaining statistics follow similar arguments. It can be decomposed

$$(T-p)^{-1} \sum_{t=p+2}^T (y_t - \hat{y}_{t|p+1})^2 = (T-p)^{-2} \sum_{t=p+2}^T (y_t - \tilde{y}_{t|p+1})^2 \quad (17)$$

$$+ T^{-2} \sum_{t=p+2}^T (\tilde{y}_{t|p+1} - \hat{y}_{t|p+1})^2 \quad (18)$$

$$T^{-2} \sum_{t=p+2}^T (y_t - \tilde{y}_{t|p+1})(\tilde{y}_{t|p+1} - \hat{y}_{t|p+1}). \quad (19)$$

Since $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p(T^{-1/2})$ and $\hat{\boldsymbol{\phi}} - \boldsymbol{\phi} = O_p(T^{-1/2})$ it holds that $\hat{y}_{t|p+1} - \tilde{y}_{t|p+1} = O_p(T^{-1/2})$ (see, for instance, Fuller 1976, p. 384) and (18) converges to zero in probability. In order to see that the term (19) also converges to zero it can be seen, without loss of generality, the case of u_t following an AR(1), $u_t = \phi u_{t-1} + a_t$. Then,

$$y_t - \tilde{y}_{t|2} = \sum_{j=3}^t \left(\sum_{i=3}^j \phi^{j-i} a_i \right) = (1 - \phi^2)^{-1} \sum_{j=3}^t a_j + \sum_{j=0}^t \alpha_j a_{t-j},$$

where $\alpha_j = \sum_{i=1}^{\infty} \phi^{j+i}$ and also $\sum_{j=0}^{\infty} |\alpha_j| < \infty$. Conversely, $\tilde{y}_{t|2} - \hat{y}_{t|2} = \sum_{j=3}^t (\phi^{j-2} - \hat{\phi}^{j-2}) u_1$, where both $\sum_{j=3}^t \phi^{j-2}$ and $\sum_{j=3}^t \hat{\phi}^{j-2}$ are bounded by an exponentially decreasing function and, therefore, $\gamma_t = \sum_{j=3}^t (\phi^{j-2} - \hat{\phi}^{j-2}) = O_p(T^{-1/2})$. Hence (19) is the average of terms that verify

$$T^{-1} \left(\frac{\gamma_t}{1 - \phi^2} \sum_{j=3}^t a_j u_1 + \gamma_t \sum_{j=0}^t \alpha_j a_{t-j} u_1 \right) \xrightarrow{p} 0$$

It is verified, then, that the term (19) converges to zero in probability. Therefore, the asymptotic distribution of EC_1^μ can be analyzed as if the parameters $\boldsymbol{\phi}$ y $\boldsymbol{\theta}$ were known.

The terms in the numerator of (17) can be decomposed as following: $y_t - \tilde{y}_{t|p+1} = \sum_{j=p+2}^t u_j - \tilde{u}_{j|p+1} = \sum_{j=p+2}^t \left(\sum_{i=p+2}^j \psi_{j-i} a_i \right)$. It can be observed that this decomposition can be written as $y_t - \tilde{y}_{t|p+1} = \psi(1)(a_{p+2} + a_{p+3} + \dots + a_t) + \nu_t$, where $\nu_t = \sum_{j=0}^{t-p-2} \alpha_j a_{t-j}$ and $\alpha_j = \sum_{i=1}^{\infty} \psi_{j+i}$. It holds that $\sum_{j=0}^{t-p-2} |\alpha_j| < \infty$. It can be written, then, $e_{t|p+1} = \psi(1) \sum_{j=1}^t a_j - \psi(1) \sum_{j=1}^{p+1} a_j + \nu_t$. Let us define the partial sums process $X_T(r)$, with $r \in [0, 1]$

$$X_T(r) = \frac{e_{[Tr]|p+1}}{T} = \psi(1) \frac{\sum_{j=1}^{[Tr]} a_j}{T} - \psi(1) \frac{\sum_{j=1}^{p+1} a_j}{T} + \frac{\nu_{[Tr]}}{T}.$$

Multiplying by $\sqrt{T}$ it can be obtained, by FCLT,

$$\psi(1) \frac{\sum_{j=1}^{[Tr]} a_j}{\sqrt{T}} \xrightarrow{d} \sigma \psi(1) W(r).$$

Also, since p is fixed,

$$\psi(1) \frac{\sum_{j=1}^{p+1} a_j}{\sqrt{T}} \xrightarrow{p} 0.$$

Since ν_t is stationary with finite variance it holds that

$$\frac{\nu_{[Tr]}}{\sqrt{T}} \xrightarrow{p} 0, \tag{20}$$

(see, for instance, Hamilton 1994, p. 482). Therefore, by the continuous mapping theorem,

$$\frac{\sum_{t=p+2}^T e_{t|p+1}^2}{T^2} \xrightarrow{d} \sigma^2 \psi(1)^2 \int_0^1 W^2(r) dr.$$

Since the denominator of PC is a consistent estimator of $\sigma^2 \psi(1)^2$ the result holds. $\square$

Critical values of B tests

Table 9: Critical values. 100,000 Monte Carlo rep.

T	Level			
	1%	2.5%	5%	10%
B_S^μ				
25	0.383	0.455	0.535	0.651
50	0.368	0.450	0.537	0.669
100	0.378	0.472	0.587	0.738
250	0.420	0.539	0.677	0.894
500	0.450	0.582	0.740	0.999
B_S^τ				
25	0.0236	0.0266	0.0298	0.0342
50	0.0213	0.0242	0.0275	0.0321
100	0.0198	0.0231	0.0264	0.0312
250	0.0192	0.0223	0.0257	0.0306
500	0.0189	0.0222	0.0256	0.0305
B_N^μ				
25	0.273	0.330	0.388	0.481
50	0.273	0.334	0.409	0.524
100	0.276	0.352	0.441	0.574
250	0.285	0.365	0.461	0.625
500	0.289	0.376	0.482	0.657
B_N^τ				
25	0.0248	0.0282	0.0318	0.0370
50	0.0229	0.0267	0.0306	0.0364
100	0.0218	0.0259	0.0298	0.0361
250	0.0214	0.0256	0.0298	0.0361
500	0.0214	0.0255	0.0298	0.0361

References

- Bhargava, A. (1986). "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies*, 53, 369–384.
- Box, G. E. P. & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- Dickey, D. A., & Fuller, W. A. (1979). "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.
- Dufour, J. M. & King, M. L. (1991). "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors," *Journal of Econometrics*, 47, 115–143.
- Elliott, G. (1997). "Efficient Tests for a Unit Root when the Initial Observation is Drawn from its Unconditional Distribution," Manuscript. University of California, San Diego.
- Elliott, G., Rothenberg, T. J. & Stock, J. H. (1992). "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.
- Ferguson, T. S. (1967). *Mathematical Statistics. A Decision Theoretic Approach*. New York. Academic Press
- Fuller, W. A. (1976), *Introduction to Statistical Time Series*. New York. John Wiley.
- Fuller, W. A. (1996), *Introduction to Statistical Time Series*. Second edition. New York. John Wiley.
- Hamilton, J. D. (1994). *Time Series Analysis*. New York. Princeton University Press.
- Hwang, J. & Schmidt, P. (1996). "Alternative methods of detrending and the power of unit root tests," *Journal of Econometrics*, 71, n227–248.
- King, M. L. (1980). "Robust Tests for Spherical Symmetry and Their Application to Least Squares Regression," *The Annals of Statistics*, 8, 1265–1271.
- King, M. L. & Hillier, G. H. (1985). "Locally Best Invariant Tests of the Error Covariance Matrix of the Linear Regression gModel," *Journal of the Royal Statistical Society, Serie B*, 47, 98–102.
- Leybourne, S. J. (1995). "Testing for Unit Roots Using Forward and Reverse Dickey–Fuller Regressions," *Oxford Bulletin of Economics and Statistics*, 57, 559–571.
- Nabeya, S. & Tanaka, K. (1990). "Limiting Power of Unit-Root Tests in Time-Series Regression," *Journal of Econometrics*, 46, 247–271.
- Pantula, S. G., Gonzalez-Farias, G. & Fuller, W. A. (1994). "A Comparison of Unit-Root Test Criteria," *Journal of Business & Economic Statistics*, 12, 449–459.
- Park, H. J. & Fuller, W. A. (1995). "Alternative Estimators and Unit Root Tests for the Autoregressive Process," *Journal of Time Series Analysis*, 16, 415–429.
- Phillips, P. C. B. (1987). "Time Series Regression With Unit Root," *Econometrica*, 55, 277–301.
- Phillips, P. C. B. & Perron P. (1988). "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.
- Said, S. E. & Dickey, D. A. (1984). "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599–607.
- Sargan, J. D. (1973). "The Durbin-Watson Ratio of the Gaussian Random Walk," . Manuscript. Department of Economics. London School of Economics. C.R.D.E.
- Sargan, J. D. & Bhargava, A. (1983). "Testing the Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk," *Econometrica*, 41, 153–174.
- Sen, D. L. & Dickey D. A. (1987). "Symmetric Test for Second Differencing in Univariate Time Series," *Journal of Business & Economic Statistics*, 5, 463–473.
- Stock, J. H. (1994). "Unit Roots, Structural Breaks, and Trends," en R. F. Engle & D. L. McFadden, eds, *Handbook of Econometrics*. Vol. IV. Elsevier. Francia.
- Tanaka, K. (1996). "Time Series Analysis. Nonstationary and Noninvertible Distribution Theory," New York. John Wiley.